Multisnapshot Sparse Bayesian Learning for DOA

Peter Gerstoft, *Member, IEEE*, Christoph F. Mecklenbräuker, *Senior Member, IEEE*,
Angeliki Xenaki, and Santosh Nannuru

***Abstract*—The directions of arrival (DOA) of plane waves are estimated from multisnapshot sensor array data using sparse Bayesian learning (SBL). The prior for the source amplitudes is assumed independent zero-mean complex Gaussian distributed with hyperparameters, the unknown variances (i.e., the source powers). For a complex Gaussian likelihood with hyperparameter, the unknown noise variance, the corresponding Gaussian posterior distribution is derived. The hyperparameters are automatically selected by maximizing the evidence and promoting sparse DOA estimates. The SBL scheme for DOA estimation is discussed and evaluated competitively against LASSO (ℓ_1-regularization), conventional beamforming, and MUSIC.**

***Index Terms*—Array processing, compressive beamforming, directions of arrival (DOA) estimation, relevance vector machine, sparse reconstruction.**

I. Introduction

Multiple measurement vector (MMV, or multiple snapshots) compressive beamforming offers several benefits over established high-resolution directions of arrival (DOA) estimators that utilize the data covariance matrix [1]–[4]: 1) it handles coherent arrivals. 2) It can be formulated with any number of snapshots in contrast to eigenvalue-based beamformers. 3) Its flexibility in formulation enables extensions to sequential processing and online algorithms [5]. 4) It achieves higher resolution than MUSIC, even in scenarios that favor these classical high-resolution methods [4].

We solve the MMV problem in the sparse Bayesian learning (SBL) framework [3] and use the maximum-*a posteriori* (MAP) estimate for DOA reconstruction. We assume complex Gaussian distributions with unknown variances (hyperparameters) for likelihood and as prior information for the source amplitudes.

The corresponding posterior distribution is also Gaussian. To determine the hyperparameters, we maximize a Type-II likelihood (evidence) for Gaussian signals hidden in Gaussian noise [6]. This has been solved with a minimization–majorization technique [7] and with expectation maximization (EM) [3], [8], [9]–[12]. Instead, we estimate the hyperparameters by iterative fitting of the unknown variances in the evidence and using stochastic maximum likelihood [13]–[15].

We propose an SBL algorithm for MMV DOA estimation which, given the number of sources, automatically estimates the set of DOAs corresponding to nonzero source power from all potential DOAs. This provides a sparse signal estimate similar to LASSO [4], [16]. Posing the problem this way, the estimated number of parameters is independent of snapshots, whereas the accuracy improves with the number of snapshots.

II. Array Data Model and Problem Formulation

Let $\boldsymbol{X}=[\boldsymbol{x}_1,\ldots,\boldsymbol{x}_L]\in\mathbb{C}^{M\times L}$ be the complex source amplitudes, x_{ml} with $m\in[1,\ldots,M]$ and $l\in[1,\ldots,L]$, at M DOAs (e.g., $\theta_m=-90^\circ+\frac{m-1}{M}180^\circ$) and L snapshots at frequency ω. We observe narrowband waves on N sensors for L snapshots $\boldsymbol{Y}=[\boldsymbol{y}_1,\ldots,\boldsymbol{y}_L]\in\mathbb{C}^{N\times L}$. A linear regression model relates the array data $\boldsymbol{Y}$ to the source amplitudes $\boldsymbol{X}$ as

$$\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{N}. \tag{1}$$

The transfer matrix $\boldsymbol{A}=[\boldsymbol{a}_1,\ldots,\boldsymbol{a}_M]\in\mathbb{C}^{N\times M}$ contains the array steering vectors for all hypothetical DOAs as columns, with the nmth element $\mathrm{e}^{-\mathrm{j}(n-1)\frac{\omega d}{c}\sin\theta_m}$ (d is the element spacing and c the sound speed). The additive noise $N\in\mathbb{C}^{N\times L}$ is assumed independent across sensors and snapshots, with each element following a complex Gaussian $\mathcal{C}N(0,\sigma^2)$.

We assume $M\gg N$ and thus (1) is underdetermined. In the presence of few stationary sources, the source vector x_l is K-sparse with $K\ll M$. We define the lth active set

$$\mathcal{M}_l=\{m\in\mathbb{N}|x_{ml}\neq 0\}=\{m_1,m_2,\ldots,m_K\} \tag{2}$$

and assume $\mathcal{M}_l=\mathcal{M}$ is constant across snapshots l. Also, we define $\boldsymbol{A}_{\mathcal{M}}\in\mathbb{C}^{N\times K}$ that contains only the K "active" columns of $\boldsymbol{A}$. $\|\cdot\|_p$ denotes the vector p-norm and $\|\cdot\|_{\mathcal{F}}$ the matrix Frobenius norm.

Similar to other DOA estimators, K can be estimated by model-order selection criteria or examining the angular spectrum. K is required only for the noise power in the proposed SBL algorithm. An inaccurate estimate influences the algorithm's convergence. Any choice $0\le K<N$ gives here a better σ^2 estimate than existing methods [3], [8]–[10], [12].

III. Bayesian Formulation

Using Bayesian inference to solve the linear problem (1) involves determining the posterior distribution of the complex source amplitudes $\boldsymbol{X}$ from the likelihood and a prior model.

A. Likelihood and Prior

Assuming the additive noise in (1) complex Gaussian, the data likelihood for the sources $\boldsymbol{X}$ given the single-frequency observations $\boldsymbol{Y}$ is complex Gaussian with noise variance σ^2 as

$$p(\boldsymbol{Y}|\boldsymbol{X};\sigma^2)=\frac{\exp\left(-\frac{1}{\sigma^2}\|\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{X}\|_{\mathcal{F}}^2\right)}{(\pi\sigma^2)^{NL}}. \tag{3}$$

Manuscript received February 25, 2016; revised July 13, 2016; accepted August 5, 2016. Date of publication August 8, 2016; date of current version September 2, 2016. This work was supported in part by the Office of Naval Research under Grant N00014-1110439 and in part by Forschungszentrum Telekommunikation Wien Austria's "Compressed channel state information feedback for time-variant MIMO channels". The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Neil Gordon.

P. Gerstoft and S. Nannuru are with the University of California San Diego, San Diego, CA 92093 USA (e-mail: gerstoft@ucsd.edu; snannuru@ucsd.edu).

C. F. Mecklenbräuker is with the Technische Universität Wien, 1040 Vienna, Austria (e-mail: cfm@nt.tuwien.ac.at).

A. Xenaki is with the Acoustic Technology, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark (e-mail: anxe@dtu.dk).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2016.2598550

For the prior, we assume that the complex source amplitudes x_{ml} are independent both across snapshots and across DOAs and follow a zero-mean complex Gaussian distribution with DOA-dependent variance $\gamma_m \in \boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_M]^T$:

$$p_m(x_{ml};\gamma_m) = \begin{cases} \delta(x_{ml}), & \text{for } \gamma_m = 0, \\ \dfrac{1}{\pi\gamma_m} \mathrm{e}^{-|x_{ml}|^2/\gamma_m}, & \text{for } \gamma_m > 0 \end{cases} \tag{4}$$

$$p(\boldsymbol{X};\boldsymbol{\gamma}) = \prod_{l=1}^{L}\prod_{m=1}^{M} p_m(x_{ml};\gamma_m) = \prod_{l=1}^{L} \mathcal{CN}(\boldsymbol{0},\boldsymbol{\Gamma}) \tag{5}$$

i.e., the source vector $\boldsymbol{x}_l$ at each snapshot $l \in [1,\ldots,L]$ is multivariate Gaussian with potentially singular covariance matrix

$$\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}) = \mathsf{E}\left[\boldsymbol{x}_l \boldsymbol{x}_l^H;\boldsymbol{\gamma}\right] \tag{6}$$

as $\mathrm{rank}(\boldsymbol{\Gamma}) = \mathrm{card}(\mathcal{M}) = K \leq M$. Note that the diagonal elements of $\boldsymbol{\Gamma}$, i.e., the hyperparameters $\boldsymbol{\gamma} \geq \boldsymbol{0}$, represent source powers. When the variance $\gamma_m = 0$, $x_{ml} = 0$ with probability 1. The sparsity of the model is thus controlled with the hyperparameters $\boldsymbol{\gamma}$, and the active set $\mathcal{M}$ is equivalently

$$\mathcal{M} = \{m \in \mathbb{N} | \gamma_m > 0\}. \tag{7}$$

The proposed SBL algorithm ultimately estimates the hyperparameters $\boldsymbol{\gamma}$ rather than the complex source amplitudes $\boldsymbol{X}$. This amounts to a significant reduction of the degrees of freedom resulting in low variance of the DOA estimates.

B. Posterior

Given the likelihood for the array observations Y (3) and the prior (5), the posterior pdf for the source amplitudes X can be found using the Bayes rule conditioned on γ, σ^2 as

$$p(\boldsymbol{X}|\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2) \equiv \frac{p(\boldsymbol{Y}|\boldsymbol{X};\sigma^2)p(\boldsymbol{X};\boldsymbol{\gamma})}{p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2)}. \tag{8}$$

The denominator $p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2)$ is the evidence term, i.e., the marginal distribution for the data, which for a given $\boldsymbol{\gamma},\sigma^2$ is a normalization factor and is neglected at first

$$p(\boldsymbol{X}|\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2) \propto p(\boldsymbol{Y}|\boldsymbol{X};\sigma^2)p(\boldsymbol{X};\boldsymbol{\gamma}) \tag{9}$$

$$\propto \frac{\mathrm{e}^{-\mathrm{tr}\left((\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})^H \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})\right)}}{(\pi^N \det \boldsymbol{\Sigma}_x)^L} = \mathcal{CN}(\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\Sigma}_x). \tag{10}$$

As both $p(\boldsymbol{Y}|\boldsymbol{X};\sigma^2)$ and $p(\boldsymbol{X};\boldsymbol{\gamma})$ in (3), (5) are Gaussians, their product (9) is Gaussian with posterior mean $\boldsymbol{\mu}_{\boldsymbol{X}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}}$:

$$\boldsymbol{\mu}_{\boldsymbol{X}} = \mathsf{E}\{X|\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2\} = \boldsymbol{\Gamma}\boldsymbol{A}^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{Y}, \tag{11}$$

$$\begin{aligned}\boldsymbol{\Sigma}_x &= \mathsf{E}\{(\boldsymbol{x}_l - \boldsymbol{\mu}_{\boldsymbol{x}_l})(\boldsymbol{x}_l - \boldsymbol{\mu}_{\boldsymbol{x}_l})^H|\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2\} \\ &= \left(\frac{1}{\sigma^2}\boldsymbol{A}^H\boldsymbol{A} + \boldsymbol{\Gamma}^{-1}\right)^{-1}\end{aligned} \tag{12}$$

where the array data covariance $\boldsymbol{\Sigma}_y$ and its inverse are derived from (1) with the matrix inversion lemma

$$\boldsymbol{\Sigma}_y = \mathsf{E}\{\boldsymbol{y}_l\boldsymbol{y}_l^H\} = \sigma^2\boldsymbol{I}_N + \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^H, \tag{13}$$

$$\boldsymbol{\Sigma}_y^{-1} = \sigma^{-2}\boldsymbol{I}_N - \sigma^{-2}\boldsymbol{A}\left(\frac{1}{\sigma^2}\boldsymbol{A}^H\boldsymbol{A} + \boldsymbol{\Gamma}^{-1}\right)^{-1}\boldsymbol{A}^H\sigma^{-2}. \tag{14}$$

If $\boldsymbol{\gamma}$ and σ^2 are known, then the MAP estimate is the posterior mean:

$$\hat{\boldsymbol{X}}^{\mathrm{MAP}} = \boldsymbol{\mu}_{\boldsymbol{X}} = \boldsymbol{\Gamma}\boldsymbol{A}^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{Y}. \tag{15}$$

The diagonal elements of $\boldsymbol{\Gamma}$ control the row-sparsity of $\hat{\boldsymbol{X}}^{\mathrm{MAP}}$ as for $\gamma_m = 0$ the corresponding mth row of $\hat{\boldsymbol{X}}^{\mathrm{MAP}}$ becomes $\boldsymbol{0}^T$.

C. Evidence

The hyperparameters $\boldsymbol{\gamma},\sigma^2$ in (11)–(14) are estimated by a type-II maximum likelihood, i.e., by maximizing the evidence that was treated as constant in (9). The evidence is the product of the likelihood (3) and the prior (5) integrated over the complex source amplitudes $\boldsymbol{X}$:

$$p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2) = \int_{\mathbb{R}^{2ML}} p(\boldsymbol{Y}|\boldsymbol{X};\sigma^2)p(\boldsymbol{X};\boldsymbol{\gamma})\,\mathrm{d}\boldsymbol{X} = \frac{\mathrm{e}^{-\mathrm{tr}\left(\boldsymbol{Y}^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{Y}\right)}}{(\pi^N \det\boldsymbol{\Sigma}_y)^L} \tag{16}$$

where $\mathrm{d}\boldsymbol{X} = \prod_l^L\prod_m^M \mathrm{Re}(\mathrm{d}X_{ml})\,\mathrm{Im}(\mathrm{d}X_{ml})$, and $\boldsymbol{\Sigma}_y$ is the data covariance (13). The L-snapshot marginal log-likelihood is

$$\log p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2) \propto -\mathrm{tr}\left(\boldsymbol{Y}^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{Y}\right) - L\log\det\boldsymbol{\Sigma}_y. \tag{17}$$

The hyperparameter estimates $\hat{\boldsymbol{\gamma}},\hat{\sigma}^2$ are obtained by maximizing the evidence:

$$(\hat{\boldsymbol{\gamma}},\ \hat{\sigma}^2) = \underset{\boldsymbol{\gamma}\geq 0,\ \sigma^2>0}{\arg\max} \log p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2). \tag{18}$$

The maximization is carried out iteratively using derivatives of the evidence for $\boldsymbol{\gamma}$ (see Section III-D) as well as conventional noise estimates (see Section III-E) as explained in Section III-F.

D. Source Power Estimation (Hyperparameters $\boldsymbol{\gamma}$)

We impose the diagonal structure $\boldsymbol{\Gamma} = \mathrm{diag}\,(\gamma)$, in agreement with (5), and form derivatives of (17) with respect to the diagonal elements γ_m, cf. [17]. Using

$$\frac{\partial\boldsymbol{\Sigma}_y^{-1}}{\partial\gamma_m} = -\boldsymbol{\Sigma}_y^{-1}\frac{\partial\boldsymbol{\Sigma}_y}{\partial\gamma_m}\boldsymbol{\Sigma}_y^{-1} = -\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m\boldsymbol{a}_m^H\boldsymbol{\Sigma}_y^{-1} \tag{19}$$

$$\frac{\partial\log\det(\boldsymbol{\Sigma}_y)}{\partial\gamma_m} = \mathrm{tr}\left(\boldsymbol{\Sigma}_y^{-1}\frac{\partial\boldsymbol{\Sigma}_y}{\partial\gamma_m}\right) = \boldsymbol{a}_m^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m \tag{20}$$

the derivative of (17) is

$$\begin{aligned}\frac{\partial\log p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2)}{\partial\gamma_m} &= \mathrm{tr}\left(\boldsymbol{Y}^H\boldsymbol{\Sigma}_y^{-1}a_m a_m^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{Y}\right) - L\boldsymbol{a}_m^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m \\ &= \|\boldsymbol{Y}^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m\|_2^2 - L\boldsymbol{a}_m^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m \\ &= \left(\frac{\gamma_m^{\mathrm{old}}}{\gamma_m^{\mathrm{new}}}\right)^2\|\boldsymbol{Y}^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m\|_2^2 - L\boldsymbol{a}_m^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m.\end{aligned} \tag{21}$$

We have introduced the factor $(\frac{\gamma_m^{\mathrm{old}}}{\gamma_m^{\mathrm{new}}})^2$ to obtain an iterative equation in γ_m. Assuming $\gamma_m, \boldsymbol{\Sigma}_y$ given (from previous iterations or initialization) and forcing (21) to zero gives the γ_m update (SBL1):

$$\gamma_m^{\mathrm{new}} = \frac{\gamma_m^{\mathrm{old}}}{\sqrt{L}}\|\boldsymbol{Y}^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m\|_2 \Big/ \sqrt{\boldsymbol{a}_m^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m}. \tag{SBL1}$$

Defining the data sample covariance matrix

$$\boldsymbol{S}_y = \boldsymbol{Y}\boldsymbol{Y}^H/L. \tag{22}$$

At the optimal solution $(\boldsymbol{\Gamma}_{\mathcal{M}}, \sigma^2)$, Jaffer's necessary condition [14, eq. (6)] must be satisfied

$$\boldsymbol{A}_{\mathcal{M}}^H\left(\boldsymbol{S}_y - \boldsymbol{\Sigma}_y\right)\boldsymbol{A}_{\mathcal{M}} = \boldsymbol{0}. \tag{23}$$

Thus, when $\boldsymbol{S}_y$ is positive definite (i.e., usually when $L \geq 2N$), we can replace $\boldsymbol{\Sigma}_y^{-1}$ in (SBL1) with $\boldsymbol{S}_y^{-1}$ [see (23)]

$$\gamma_m^{\text{new}} = \frac{\gamma_m^{\text{old}}}{\sqrt{L}}\|\boldsymbol{Y}^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m\|_2\Big/\sqrt{\boldsymbol{a}_m^H\boldsymbol{S}_y^{-1}\boldsymbol{a}_m}. \tag{SBL}$$

The SBL estimate tends to converge faster as the denominator does not change during iterations.

Wipf and Rao [3] [see Eq. (18)] followed a less direct EM approach to estimate the update M-SBL:

$$\gamma_m^{\text{new}} = \frac{(\gamma_m^{\text{old}})^2}{L}\|\boldsymbol{Y}^H\boldsymbol{\Sigma}_y^{-1}\boldsymbol{a}_m\|_2^2 + (\boldsymbol{\Sigma}_x)_{mm}. \tag{M-SBL}$$

The sequence of parameter estimates in the EM iteration converges [18]. However, convergence is only guaranteed toward a *local* optimum of the marginal log-likelihood (17). As seen from simulations (see Section IV), all the update rules (SBL1)–(M-SBL) converge. Guarantees for convergence can be given provided $|\partial\gamma_m^{\text{new}}/\partial\gamma_m| < 1$ but is hard to prove analytically.

E. Noise Variance Estimation (Hyperparameter σ^2)

Obtaining a good noise variance estimate is important for fast convergence of the SBL method, as it controls the sharpness of the peaks. For a given set of active DOAs $\mathcal{M}$, stochastic maximum likelihood [11], [13], [15] provides an asymptotically efficient estimate of σ^2 similar to our (27).

Let $\boldsymbol{\Gamma}_{\mathcal{M}} = \text{diag}(\boldsymbol{\gamma}_{\mathcal{M}}^{\text{new}})$ be the covariance matrix of the K active sources obtained above with corresponding active steering matrix $\boldsymbol{A}_{\mathcal{M}}$ that maximizes (17). The corresponding data covariance matrix is

$$\boldsymbol{\Sigma}_y = \sigma^2\boldsymbol{I}_N + \boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H \tag{24}$$

where $\boldsymbol{I}_N$ is the identity matrix of order N. The data covariance models (13) and (24) are identical. Substituting (24) into (23) gives

$$\boldsymbol{A}_{\mathcal{M}}^H\left(\boldsymbol{S}_y - \sigma^2\boldsymbol{I}_N\right)\boldsymbol{A}_{\mathcal{M}} = \boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}}. \tag{25}$$

Defining the projection matrix onto the subspace spanned by the active steering vectors $\boldsymbol{P} = \boldsymbol{A}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^+ = \boldsymbol{A}_{\mathcal{M}}(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{A}_{\mathcal{M}}^H = \boldsymbol{P}^H = \boldsymbol{P}^2$. Equation (25) is left multiplied with $\boldsymbol{A}_{\mathcal{M}}^{+H} = \boldsymbol{A}_{\mathcal{M}}(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}})^{-1}$ and right multiplied with $\boldsymbol{A}_{\mathcal{M}}^+ = (\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{A}_{\mathcal{M}}^H$. Then

$$\begin{aligned}\boldsymbol{P}\boldsymbol{S}_y\boldsymbol{P}^H - \sigma^2\boldsymbol{P}\boldsymbol{P}^H &= \boldsymbol{P}\boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{P}^H\\ &= \boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H = \boldsymbol{\Sigma}_y - \sigma^2\boldsymbol{I}_N \end{aligned} \tag{26}$$

Evaluating the trace and defining $\epsilon = \text{tr}[\boldsymbol{\Sigma}_y - \boldsymbol{S}_y]$ gives

$$\sigma^2 = \frac{\text{tr}[(\boldsymbol{I}_N - \boldsymbol{P})\boldsymbol{S}_y] + \epsilon}{N - K} \approx \frac{\text{tr}[(\boldsymbol{I}_N - \boldsymbol{P})\boldsymbol{S}_y]}{N - K}. \tag{27}$$

The noise power estimate (27) is *error-free* if $\text{tr}[\boldsymbol{\Sigma}_y] = \text{tr}[\boldsymbol{S}_y]$, *unbiased* because $\text{E}[\epsilon] = 0$, *consistent* since its variance also tends to zero for $L \to \infty$ [19], and *asymptotically efficient* as it approaches the CRLB for $L \to \infty$ [20].

TABLE I
SBL ALGORITHM

Initialize, here: $\sigma^2 = 0.1, \boldsymbol{\gamma} = 1, \epsilon_{\min} = 0.001, j_{\max} = 500$	
while $(\epsilon > \epsilon_{\min})$ and $(j < j_{\max})$	
$\quad j = j + 1, \quad \boldsymbol{\gamma}^{\text{old}} = \boldsymbol{\gamma}^{\text{new}}, \quad \boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma}^{\text{new}})$	
$\quad \boldsymbol{\Sigma}_y = \sigma^2\boldsymbol{I}_N + \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^H$	(13)
$\quad \gamma_m^{\text{new}} =$ choose Eq. (SBL1), (SBL), or (M-SBL)	
$\quad \mathcal{M} = \{m \in \mathbb{N} \mid$ K largest peaks in $\boldsymbol{\gamma}\} = \{m_1 \dots m_K\}$	(7)
$\quad \boldsymbol{A}_{\mathcal{M}} = (\boldsymbol{a}_{m_1}, \dots, \boldsymbol{a}_{m_K})$	
$\quad (\sigma^2)^{\text{new}} = \frac{1}{N-K}\,\text{tr}\left((\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^+)\boldsymbol{S}_y\right)$	(27)
$\quad \epsilon = \|\boldsymbol{\gamma}^{\text{new}} - \boldsymbol{\gamma}^{\text{old}}\|_1 / \|\boldsymbol{\gamma}^{\text{old}}\|_1$	(29)
Output: $\mathcal{M}, \boldsymbol{\gamma}^{\text{new}}, (\sigma^2)^{\text{new}}$	

Several estimators for the noise σ^2 are proposed based on EM [3], [9], [10], [21], [22]. For a comparative illustration in Section IV, we use the iterative noise σ^2 EM estimate in [22]

$$(\sigma^2)^{\text{new}} = \frac{\frac{1}{L}\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{\mu}_{\boldsymbol{X}}\|_{\mathcal{F}}^2 + (\sigma^2)^{\text{old}}\left(M - \sum_{i=1}^{M}\frac{(\Sigma_x)_{ii}}{\gamma_i}\right)}{N}. \tag{28}$$

Empirically, EM noise estimates as (28) converge to zero in our application.

F. SBL Algorithm

The algorithm is summarized in Table I. Given the observed $\boldsymbol{Y}$, we iteratively update $\boldsymbol{\Sigma}_y$ (13) by using the current $\boldsymbol{\gamma}$. $\boldsymbol{\Sigma}_y^{-1}$ is computed directly as inverse of $\boldsymbol{\Sigma}_y$ as that is more efficient than using (14). Either SBL, SBL1, or M-SBL can update γ_m for $m = 1, \dots, M$ and then (27) is used to estimate σ^2. The convergence rate ϵ measures the relative change in estimated total source power:

$$\epsilon = \|\boldsymbol{\gamma}^{\text{new}} - \boldsymbol{\gamma}^{\text{old}}\|_1 \;/\; \|\boldsymbol{\gamma}^{\text{old}}\|_1. \tag{29}$$

The algorithm stops when $\epsilon \leq \epsilon_{\min}$ and the output is the active set $\mathcal{M}$ (7) from which all source parameter are computed.

IV. EXAMPLE

Compressive beamforming achieves high-resolution and reliable DOA estimation even with a single snapshot [1], [2], [5], [23]–[25]. The identifiability for SBL has been addressed previously [26], [27]. For multiple sources with well-separated DOAs and similar magnitudes, conventional beamforming (CBF) and LASSO/SBL methods provide similar DOA estimates. They differ, however, in their behavior whenever two sources are closely spaced.

Thus, we examine a scenario with three sources at DOAs $[-3,\ 2,\ 75]^\circ$ with magnitudes $[12,\ 22,\ 20]$ dB [4]. We consider an array with $N = 20$ elements, half-wavelength spacing, and $L = 50$ snapshots are observed. The DOAs are on an angular grid $[-90{:}0.5 : 90]^\circ$, $M = 361$. The noise is modeled as iid complex Gaussian, though robustness to array imperfections [28] and extreme noise distributions [29] can be important. The single-snapshot array signal-to-noise ratio (SNR) is $\text{SNR} = 10\log_{10}[\mathsf{E}\left\{\|\boldsymbol{A}\boldsymbol{x}_l\|_2^2\right\}/\mathsf{E}\left\{\|\boldsymbol{n}_l\|_2^2\right\}]$.

Then, for L snapshots the noise power σ_T^2 is

$$\sigma_T^2 = \mathsf{E}[\|\boldsymbol{N}\|_{\mathcal{F}}^2]/L/N = 10^{-\text{SNR}/10}\ \mathsf{E}\,\|\boldsymbol{A}\boldsymbol{X}\|_{\mathcal{F}}^2/L/N. \tag{30}$$

The estimated $(\sigma^2)^{\text{new}}$ (27) deviates from σ_T^2 (30) randomly.

The performance of several DOA estimation methods is compared in Fig. 1. The LASSO solution is based on multiple snapshots [4] and programmed in CVX [30], [31]. SBL and

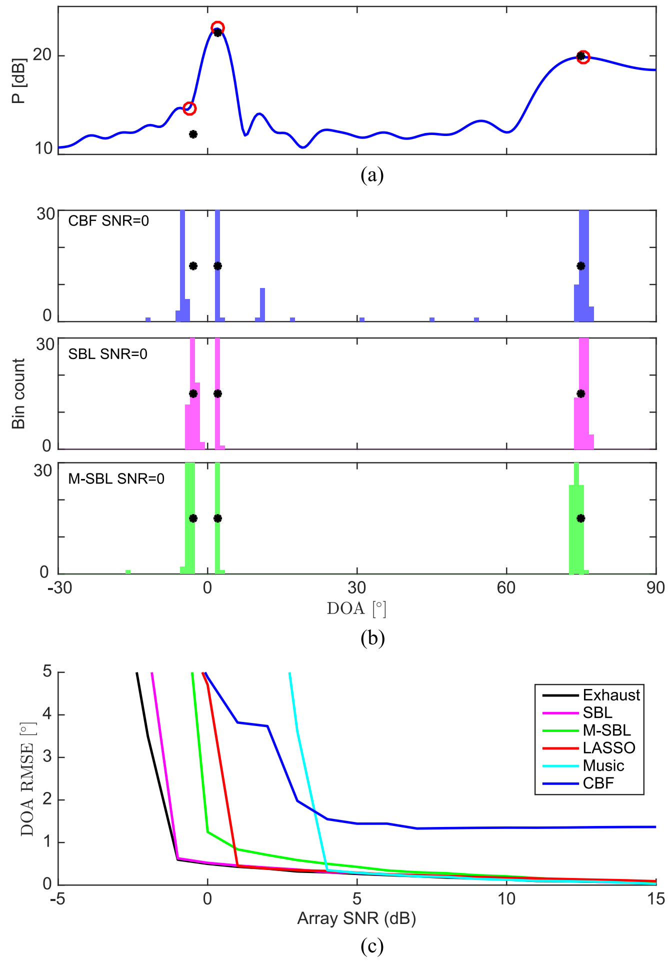

Fig. 1. Multiple $L = 50$ snapshot example for sources at DOAs $[-3,\ 2,\ 75]^\circ$. (a) Spectra for CBF and SBL (o) at SNR $= 0$ dB. (b) CBF, SBL, and M-SBL histogram based on 100 Monte Carlo simulations at SNR $= 0$ dB. (c) RMSE performance versus array SNR for exhaustive, SBL, M-SBL, LASSO, MUSIC, and CBF. The true source positions (●) are indicated in (a) and (b).

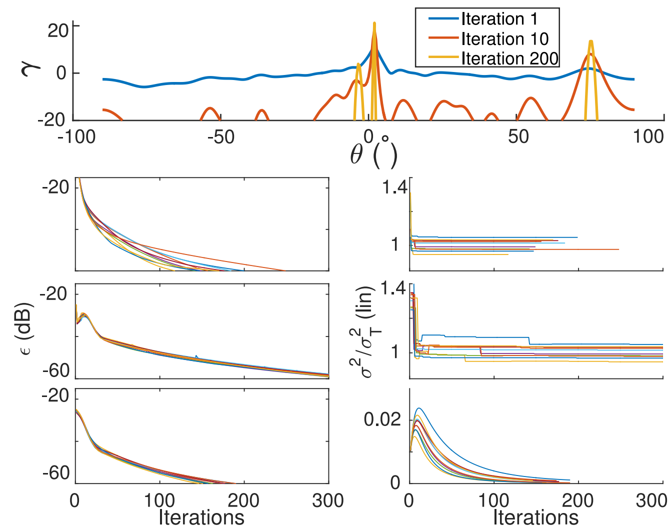

Fig. 2. Convergence at SNR $= 0$ dB with $L = 50$. (1st row) $\boldsymbol{\gamma}$ at iteration 1, 10, 200 for SBL. Convergence of (2nd row) SBL and (3rd and 4th row) M-SBL for 10 Monte Carlo simulations. Convergence is shown for ϵ (left) and σ^2/σ_T^2 (right). In (2nd and 3rd row) the noise estimate $(\sigma^2)^{\text{new}}$ is based on (27) and in (4th row) (28).

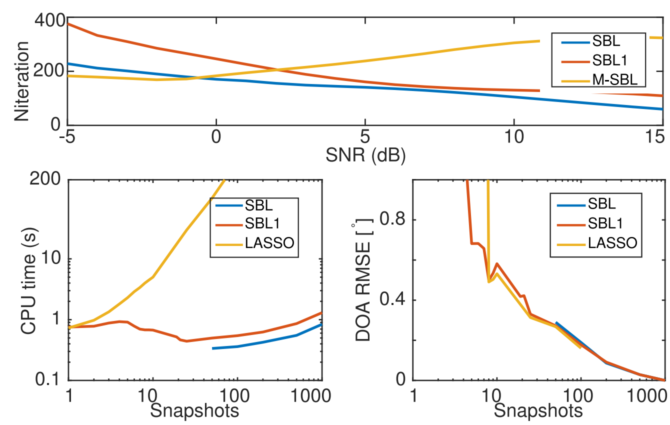

Fig. 3. (a) Average number of iterations at each SNR for M-SBL, SBL1, and SBL with $L = 50$ snapshots. At array SNR $= 5$ dB, (b) average CPU time (Macbook Pro 2014), and (c) RMSE for LASSO and SBL versus the number of snapshots. All results are an average of 100 Monte Carlo simulations.

M-SBL are calculated using the pseudocode in Table I. At array SNR $= 0$ dB, the histograms [see Fig. 1(b)] show that SBL and M-SBL localize the sources well, in contrast to CBF. The root mean squared error (RMSE) in Fig. 1(c) shows low resolution of CBF due to the broad main lobe [see Fig. 1(a)] and MUSIC performs well for $\mathrm{SNR} > 5\,\mathrm{dB}$. We include exhaustive search, which gives the exact conditional ML estimate by globally optimizing (3) for all DOA combinations. It requires $361!/(3!358!) = 7.8 \cdot 10^6$ evaluations. LASSO and the SBL methods perform better than MUSIC and offer similar accuracy to exhaustive search.

The spatial spectrum [see Fig. 2(1st row)] shows that $\boldsymbol{\gamma}$ improves with SBL iterations from initially locating only the main peak to locating also the weaker sources. SBL exhibits faster convergence than M-SBL to $\epsilon_{\min} = -60$ dB where the algorithm stops [see Fig. 2(2nd row) versus (3rd row)]. M-SBL underestimates σ^2 significantly when using (28) [see Fig. 2(4th row)].

The average number of iterations for SBL and SBL1 decreases with SNR but increases for M-SBL [see Fig. 3(a)]. As the number of snapshots increases, the RMSE error goes to zero because the sources are on the grid [see Fig. 3(c)]. For SBL and SBL1, the CPU time is nearly constant with the number of snapshots [see Fig. 3(b)]. The number of estimated parameters $(\boldsymbol{\gamma}, \sigma^2)$ is independent of the number of snapshots, whereas the RMSE decreases with the growing number of snapshot. Contrarily, for LASSO the number of degrees of freedom in $\boldsymbol{X}$ increases with the number of snapshots.

V. Conclusion

An SBL algorithm is derived for high-resolution DOA estimation from multisnapshot complex-valued array data. The algorithm uses evidence maximization based on derivatives to estimate the source powers and the noise variance. The method uses the estimated source power at each potential DOA as a proxy for an active DOA promoting sparse reconstruction.

Simulations indicate that the proposed SBL algorithm is a factor of 2 faster than established EM approaches at the same estimation accuracy. Increasing the number of snapshots improves the estimation accuracy, whereas the computational effort is nearly independent of snapshots.

REFERENCES

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

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

[3] 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.

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

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

[6] 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.

[7] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Process.*, vol. 92, no. 7, pp. 1580–1590, 2012.

[8] D. P. Wipf and S. Nagarajan, "Beamforming using the relevance vector machine," in *Proc. 24th Int. Conf. Mach. Learn.*, New York, NY, USA, 2007, pp. 1–8.

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

[10] Z. Zhang and B. D Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sep. 2011.

[11] 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.

[12] J. A. Zhang, Z. Chen, P. Cheng, and X. Huang, "Multiple-measurement vector based implementation for single-measurement vector sparse Bayesian learning with reduced complexity," *Signal Process.*, vol. 118, pp. 153–158, 2016.

[13] J. F. Böhme, "Source-parameter estimation by approximate maximum likelihood and nonlinear regression," *IEEE J. Ocean. Eng.*, vol. JOE-10, no. 3, pp. 206–212, Jul. 1985.

[14] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: A separable solution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, 1988, pp. 2893–2896.

[15] P. Stoica and A. Nehorai, "On the concentrated stochastic likelihood function in array processing," *Circuits Syst. Signal Process.*, vol. 14, no. 5, pp. 669–674, 1995.

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

[17] J. F. Böhme, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Process.*, vol. 11, pp. 329–337, 1986.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B*, vol. 39, pp. 1–38, 1977.

[19] D. Kraus, *Approximative Maximum-Likelihood-Schätzung und verwandte Verfahren zur Ortung und Signalschätzung mit Sensorgruppen*. Aachen, Germany: Shaker Verlag, 1993.

[20] G. Bienvenu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 5, pp. 1235–1248, Oct. 1983.

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

[22] Z. Zhang, T.-P. Jung, S. Makeig, P. Zhouyue, and B. D. Rao, "Spatiotemporal sparse Bayesian learning with applications to compressed sensing of multichannel physiological signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 6, pp. 1186–1197, Nov. 2014.

[23] G. F. Edelmann and C. F. Gaumond, "Beamforming using compressive sensing," *J. Acoust. Soc. Amer.*, vol. 130, no. 4, pp. 232–237, 2011.

[24] S. Fortunati, R. Grasso, F. Gini, M. S. Greco, and K. LePage, "Single-snapshot DOA estimation by using compressed sensing," *EURASIP J. Advances Signal Process.*, vol. 120, no. 1, pp. 1–17, 2014.

[25] A. Xenaki and P. Gerstoft, "Grid-free compressive beamforming," *J. Acoust. Soc. Amer.*, vol. 137, no. 1923–1935, 2015.

[26] R. Prasad and C. R. Murthy, "Cramer-rao-type bounds for sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 622–632, Feb. 2013.

[27] P. Pal and P. P. Vaidyanathan, "Parameter identifiability in sparse Bayesian learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2014, pp. 1851–1855.

[28] C. Weiss and A. M. Zoubir, "DOA estimation in the presence of array imperfections: A sparse regularization parameter selection problem," in *Proc. IEEE Workshop Statist. Signal Process.*, Jun. 2014, pp. 348–351.

[29] E. Ollila, "Multichannel sparse recovery of complex-valued signals using Huber's criterion," in *Proc. 3_{rd} Int. Workshop Compressed Sens. Theory Appl. Radar, Sonar, Remote Sens.*, Pisa, Italy, Jun. 2015, pp. 1–4.

[30] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1. (2016). [Online]. Available: cvxr.com/cvx. Accessed on: Feb. 9, 2016.

[31] C. F. Mecklenbräuker, P. Gerstoft, and E. Zöchmann, "c–LASSO and its dual for sparse signal estimation from array data," *Signal Process.*, Jul. 1, 2016, ISSN 0165-1684, Available [online]: http://dx.doi.org/10.1016/j.sigpro.2016.06.029.