WIDEBAND SPARSE BAYESIAN LEARNING FOR DOA ESTIMATION FROM MULTIPLE SNAPSHOTS

Peter Gerstoft *

Marine Physics Lab
University of California San Diego
La Jolla, CA, USA.

Christoph F. Mecklenbräuker †

Christian Doppler Laboratory
Wireless technologies for sustainable mobility
Inst. of Telecommunications, TU Wien, Austria

ABSTRACT

The directions of arrival (DOA) of plane waves are estimated from multi-frequency multi-snapshot sensor array data using Sparse Bayesian Learning (SBL). The prior for the source amplitudes is assumed to be independently zero-mean complex Gaussian distributed with hyperparameters being the unknown variances (i.e. the source powers). For a complex Gaussian likelihood with unknown noise variance hyperparameter, the corresponding Gaussian posterior distribution is derived. For a given number of DOAs, the hyperparameters are automatically selected by maximizing the evidence and promote sparse DOA estimates. The SBL scheme for DOA estimation is discussed and evaluated competitively against MUSIC.

1. INTRODUCTION

In direction of arrival (DOA) estimation based on single snapshot data, compressive beamforming, i.e. sparse processing, achieves high-resolution DOA estimation and acoustic imaging [1, 2, 3, 4, 5, 6], outperforming traditional methods[7].

Multiple measurement vector (MMV, or multiple snapshots) compressive beamforming offers several benefits over established high-resolution DOA estimators based on the data covariance[1, 5, 8, 9]. It can be formulated for 1) partially coherent arrivals, 2) any number of snapshots, and 3) extensions to sequential processing, and online algorithms [3]. It achieves higher resolution than MUSIC, even in scenarios that favor these classical high-resolution methods [9].

We solve the MMV problem in the sparse Bayesian learning (SBL) framework[8] and use the maximum-a-posteriori (MAP) estimate for DOA reconstruction. We assume complex Gaussian distributions with unknown variances (hyperparameters) both for the likelihood and as prior information for the source amplitudes. Hence, 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. This has been solved with a Minimization-majorization technique[10] and with expectation maximization (EM) [8, 11, 12, 13]. Instead, we estimate the hyperparameters directly from the likelihood derivatives using stochastic maximum likelihood[14, 15, 16].

We propose a SBL algorithm for MMV DOA estimation which, given the number of sources, automatically estimates the set of DOAs corresponding to non-zero source power from all potential DOAs. This provides a sparse signal estimate similar to LASSO[17, 9]. Posing the problem this way, the estimated number of parameters is independent of snapshots, while the accuracy improves with the number of snapshots.

1.1. Array data model

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

$$\boldsymbol{Y}_f = \boldsymbol{A}_f \boldsymbol{X}_f + \boldsymbol{N}_f \ , \qquad \forall f. \tag{1}$$

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

We assume $M \gg N$ and thus (1) is underdetermined. In the presence of few stationary sources, the source vector $\boldsymbol{x}_{lf}$ is K-sparse with $K \ll M$. We define the active set for snapshot l and frequency f as

$$\mathcal{M}_{lf} = \{m \in \mathbb{N} | x_{mlf} \neq 0\} = \{m_1, m_2, \dots, m_K\} \ , \tag{2}$$

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

*Supported by the Office of Naval Research, Grant Nos. N00014-1110439 and N00014-1310632 (MURI)

†Supported by CD Lab Wireless technologies for sustainable mobility

2. 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.

2.1. Likelihood and Prior

Assuming the additive noise (1) complex Gaussian the data likelihood, i.e., the conditional probability density function (pdf) for the multi-frequency observation $\boldsymbol{Y} = [\boldsymbol{Y}_1^T \ldots \boldsymbol{Y}_F^T]^T$ given the sources $\boldsymbol{X} = [\boldsymbol{X}_1^T \ldots \boldsymbol{X}_F^T]^T$ is complex Gaussian with noise variance σ_f^2.

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

The complex source amplitudes x_{mlf} are assumed independent across snapshots l, across DOAs m, and frequency bins f, They follow a zero-mean complex Gaussian distribution with DOA– and frequency dependent variance γ_{mf}. $\boldsymbol{\gamma}_f = [\gamma_{1f}, \ldots, \gamma_{Mf}]^T$ and $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1 \ldots \boldsymbol{\gamma}_F]$,

$$p_m(x_{mlf};\gamma_{mf}) = \begin{cases} \delta(x_{mlf}) & \text{for } \gamma_{mf}=0 \\ \frac{1}{\pi\gamma_{mf}} \mathrm{e}^{-|x_{mlf}|^2/\gamma_{mf}} & \text{for } \gamma_{mf}>0 \end{cases}$$

$$p(\boldsymbol{X};\boldsymbol{\gamma}) = \prod_{m,l,f} p_m(x_{mlf};\gamma_{mf}) = \prod_{f=1}^{F}\prod_{l=1}^{L} \mathcal{CN}(\boldsymbol{0},\boldsymbol{\Gamma}_f), \quad (4)$$

i.e., the source vector $\boldsymbol{x}_{lf}$ at each snapshot l and frequency bin f has a multivariate Gaussian distribution with potentially singular covariance matrix,

$$\boldsymbol{\Gamma}_f = \operatorname{diag}(\boldsymbol{\gamma}_f) = \mathsf{E}\left[\boldsymbol{x}_{lf}\boldsymbol{x}_{lf}^H;\boldsymbol{\gamma}_f\right], \quad (5)$$

as $\operatorname{rank}(\boldsymbol{\Gamma}_f) = \operatorname{card}(\mathcal{M}) = K \leq M$. Note that the diagonal elements of $\boldsymbol{\Gamma}_f$, i.e., the hyperparameters $\boldsymbol{\gamma}_f \geq \boldsymbol{0}$, represent source powers. When the variance $\gamma_{mf} = 0$, then $x_{mlf} = 0$ with probability 1. The sparsity of the model is thus controlled with the hyperparameters $\boldsymbol{\gamma}$.

2.2. Posterior

Given the likelihood for the array observations $\boldsymbol{Y}$ (3) and the prior (4), the posterior pdf for the source amplitudes $\boldsymbol{X}$ can be found using Bayes rule conditioned on $\boldsymbol{\gamma}, \sigma^2$,

$$p(\boldsymbol{X}|\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2) \equiv \frac{p(\boldsymbol{Y}|\boldsymbol{X};\sigma^2)p(\boldsymbol{X};\boldsymbol{\gamma})}{p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2)}. \quad (6)$$

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,

$$p(\boldsymbol{X}|\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2) \propto p(\boldsymbol{Y}|\boldsymbol{X};\sigma^2)p(\boldsymbol{X};\boldsymbol{\gamma}) \quad (7)$$

$$\propto \frac{\mathrm{e}^{-\operatorname{tr}\left((\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})^H\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}})\right)}}{(\pi^N \det \boldsymbol{\Sigma}_x)^L} = \mathcal{CN}(\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{x}}). \quad (8)$$

As both $p(\boldsymbol{Y}|\boldsymbol{X};\sigma^2)$ in (3) and $p(\boldsymbol{X};\boldsymbol{\gamma})$ in (4) are Gaussians, their product (7) is Gaussian with posterior mean $\boldsymbol{\mu}_{\boldsymbol{X}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}}$,

$$\boldsymbol{\mu}_{\boldsymbol{X}} = \mathsf{E}\{\boldsymbol{X}|\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2\} = \boldsymbol{\Gamma}\boldsymbol{A}^H\Sigma_{\boldsymbol{y}}^{-1}\boldsymbol{Y}, \quad (9)$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{x}} &= \mathsf{E}\{(\boldsymbol{x}_l - \boldsymbol{\mu}_{\boldsymbol{x}_l})(\boldsymbol{x}_l - \boldsymbol{\mu}_{\boldsymbol{x}_l})^H|\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2\} \\ &= \left(\frac{1}{\sigma^2}\boldsymbol{A}^H\boldsymbol{A} + \boldsymbol{\Gamma}^{-1}\right)^{-1} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{A}^H\Sigma_{\boldsymbol{y}}^{-1}\boldsymbol{A}\boldsymbol{\Gamma}, \quad (10)\end{aligned}$$

where $\boldsymbol{\Gamma} = \operatorname{diag}(\boldsymbol{\gamma})$, $\boldsymbol{A} = \operatorname{diag}(\boldsymbol{A}_1 \ldots \boldsymbol{A}_F)$, the array data covariance is $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ and its inverse are derived from (1) using the matrix inversion lemma

$$\boldsymbol{\Sigma}_{\boldsymbol{y}} = \mathsf{E}\{\boldsymbol{y}_l\boldsymbol{y}_l^H\} = \sigma^2\boldsymbol{I}_{NF} + \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^H, \quad (11)$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1} &= \sigma^{-2}\boldsymbol{I}_{NF} - \sigma^{-2}\boldsymbol{A}\left(\frac{1}{\sigma^2}\boldsymbol{A}^H\boldsymbol{A} + \boldsymbol{\Gamma}^{-1}\right)^{-1}\boldsymbol{A}^H\sigma^{-2} \\ &= \sigma^{-2}\boldsymbol{I}_{NF} - \sigma^{-2}\boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{A}^H\sigma^{-2}. \quad (12)\end{aligned}$$

For $\boldsymbol{\gamma}$ and σ^2 known, the MAP estimate is the posterior mean,

$$\hat{\boldsymbol{X}}^{\mathrm{MAP}} = \boldsymbol{\mu}_{\boldsymbol{X}} = \boldsymbol{\Gamma}\boldsymbol{A}^H\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{Y}\ . \quad (13)$$

The elements of $\boldsymbol{\gamma}$ control the row-sparsity of $\hat{\boldsymbol{X}}^{\mathrm{MAP}}$ as for γ_{mf}=0 the corresponding row of $\hat{\boldsymbol{X}}^{\mathrm{MAP}}$ is $\boldsymbol{0}^T$.

For wideband signals we assume that the signal direction is frequency independent. Since γ_{mf} represent the variance, we form a composite $\tilde{\gamma}_m = \sum_{f=1}^{F}\gamma_{mf}$. Thus, the active set $\mathcal{M}$ is equivalently defined by

$$\mathcal{M} = \{m \in \mathbb{N}|\tilde{\gamma}_m = \textstyle\sum_{f=1}^{F}\gamma_{mf} > 0\}\ . \quad (14)$$

2.3. Evidence

The hyperparameters $\boldsymbol{\gamma}, \sigma^2$ in (9–12) are estimated by a type-II maximum likelihood, i.e., by maximizing the evidence which was treated as constant in (7). The evidence is the product of the likelihood (3) and the prior (4) integrated over the complex source amplitudes $\boldsymbol{X}$,

$$p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2) = \int_{\mathbb{R}^{2ML}} p(\boldsymbol{Y}|\boldsymbol{X};\sigma^2)p(\boldsymbol{X};\boldsymbol{\gamma})\,\mathrm{d}\boldsymbol{X} = \frac{\mathrm{e}^{-\operatorname{tr}\left(\boldsymbol{Y}^H\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{Y}\right)}}{(\pi^N \det \boldsymbol{\Sigma}_{\boldsymbol{y}})^L}, \quad (15)$$

where $\mathrm{d}\boldsymbol{X} = \prod_{l=1}^{L}\prod_{m=1}^{M}\operatorname{Re}(\mathrm{d}X_{ml})\operatorname{Im}(\mathrm{d}X_{ml})$, and $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ is the data covariance (11). The L-snapshot marginal log-likelihood becomes

$$\begin{aligned}\log p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2) &\propto -\operatorname{tr}\left(\boldsymbol{Y}^H\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{Y}\right) - L\log\det\boldsymbol{\Sigma}_{\boldsymbol{y}} \\ &\propto -\operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{S}_{\boldsymbol{y}}\right) - \log\det\boldsymbol{\Sigma}_{\boldsymbol{y}}, \quad (16)\end{aligned}$$

where we define the data sample covariance matrix,

$$\boldsymbol{S}_{\boldsymbol{y}} = \boldsymbol{Y}\boldsymbol{Y}^H/L. \quad (17)$$

Note that (16) does not involve the inverse of $\boldsymbol{S_y}$ hence it works well even for few snapshots (small L).

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 Sec. 2.4) as well as conventional noise estimates (see Sec. 2.5) as explained in Sec. 2.6.

2.4. Source power estimation (hyperparameters $\boldsymbol{\gamma}$)

We impose the diagonal structure $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$, in agreement with (4), and form derivatives of (16) with respect to the diagonal elements γ_{mf}, cf. [18]. Using

$$\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{y}_f}^{-1}}{\partial \gamma_{mf}} = -\boldsymbol{\Sigma}_{\boldsymbol{y}_f}^{-1}\boldsymbol{a}_{mf}\boldsymbol{a}_{mf}^H\boldsymbol{\Sigma}_{\boldsymbol{y}_f}^{-1}, \tag{19}$$

$$\frac{\partial \log\det(\boldsymbol{\Sigma}_{\boldsymbol{y}_f})}{\partial \gamma_{mf}} = \boldsymbol{a}_{mf}^H\boldsymbol{\Sigma}_{\boldsymbol{y}_f}^{-1}\boldsymbol{a}_{mf}, \tag{20}$$

the derivative of (16) is

$$\frac{\partial \log p(\boldsymbol{Y};\boldsymbol{\gamma},\sigma^2)}{\partial \gamma_m} = \frac{1}{\gamma_{mf}^2 L}\|\boldsymbol{\mu}_{mf}\|_2^2 - \boldsymbol{a}_{mf}^H\boldsymbol{\Sigma}_{\boldsymbol{y}_f}^{-1}\boldsymbol{a}_{mf}, \tag{21}$$

where $\boldsymbol{\mu}_{mf} = \gamma_{mf}\boldsymbol{a}_{mf}^H\boldsymbol{\Sigma}_{\boldsymbol{y}_f}^{-1}\boldsymbol{Y}_f$ is the $[(f-1)M+m]$th row of $\boldsymbol{\mu}_{\boldsymbol{X}}$ in (9). As $\boldsymbol{\mu}_{mf}$ is given (from earlier iterations or initialization), forcing (21) to zero gives the γ_{mf} update:

$$\gamma_{mf}^{\mathrm{new}} = \frac{1}{\sqrt{L}}\|\boldsymbol{\mu}_{mf}\|_2 / \sqrt{\boldsymbol{a}_{mf}^H\boldsymbol{\Sigma}_{\boldsymbol{y}_f}^{-1}\boldsymbol{a}_{mf}}. \tag{22}$$

When the sample data covariance $\boldsymbol{S_y}$ is positive definite (i.e. usally when $L \geq 2N$) we can replace $\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}$ in (22) with $\boldsymbol{S}_{\boldsymbol{y}}^{-1}$ [see (26)]

$$\gamma_{mf}^{\mathrm{new}} = \frac{1}{\sqrt{L}}\|\boldsymbol{\mu}_{mf}\|_2 / \sqrt{\boldsymbol{a}_{mf}^H\boldsymbol{S}_{\boldsymbol{y}_f}^{-1}\boldsymbol{a}_{mf}}. \tag{23}$$

The 23 estimate tends to converge faster as the denominator does not change during iterations.

Wipf and Rao ([8]: Eq.(18)) followed the EM approach to estimate the update 24:

$$\gamma_{mf}^{\mathrm{new}} = \frac{1}{L}\|\boldsymbol{\mu}_{mf}\|_2^2 + [(\boldsymbol{\Sigma}_{\boldsymbol{x}})_{mm}]_f. \tag{24}$$

The sequence of parameter estimates in the EM iteration has been proven to converge [19]. However, the convergence is only guaranteed towards a *local* optimum of the marginal log-likelihood (16). As shown in Sec. 3 all the update rules (22)–(24) converge provided $|\partial\gamma_{mf}^{\mathrm{new}}/\partial\gamma_{mf}| < 1$.

0	Given: $\boldsymbol{A} \in \mathbb{C}^{NF\times MF}$ $\boldsymbol{Y} \in \mathbb{C}^{N\times L}$, $K=3$ Init: $\sigma_0^2=0.1, \gamma_0=1, \epsilon_{\min}=0.001, j_{\max}=500$	
1	initialize $j=0, \sigma^2=\sigma_0^2, \boldsymbol{\gamma}=\gamma_0$	
2	while $(\epsilon > \epsilon_{\min})$ and $(j < j_{\max})$	
3	$j=j+1,\ \boldsymbol{\gamma}^{\mathrm{old}}=\boldsymbol{\gamma}^{\mathrm{new}},\ \boldsymbol{\Gamma}=\mathrm{diag}(\boldsymbol{\gamma}^{\mathrm{new}})$	
4	$\boldsymbol{\Sigma_y} = \sigma^2\boldsymbol{I}_{NF} + \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^H$	(11)
5	$\forall m,f:\ \boldsymbol{\mu}_{mf} = \gamma_{mf}\boldsymbol{a}_{mf}^H\boldsymbol{\Sigma}_{\boldsymbol{y}_f}^{-1}\boldsymbol{Y}_f$	(9)
6	$\gamma_{mf}^{\mathrm{new}} = \begin{cases} \frac{1}{\sqrt{L}}\|\boldsymbol{\mu}_{mf}\|_2 \big/ \sqrt{\boldsymbol{a}_{mf}^H\boldsymbol{S}_{\boldsymbol{y}_f}^{-1}\boldsymbol{a}_{mf}} \\ \frac{1}{\sqrt{L}}\|\boldsymbol{\mu}_{mf}\|_2 \big/ \sqrt{\boldsymbol{a}_{mf}^H\boldsymbol{\Sigma}_{\boldsymbol{y}_f}^{-1}\boldsymbol{a}_{mf}} \\ \frac{1}{L}\|\boldsymbol{\mu}_{mf}\|_2^2 + [(\boldsymbol{\Sigma}_{\boldsymbol{x}})_{mm}]_f \end{cases}$	(23) (22) (24)
7	$\tilde{\gamma}_m = \sum_{f=1}^F \gamma_{mf}$ (wideband estimate) $\mathcal{M}=\{K$ highest peaks in $\tilde{\boldsymbol{\gamma}}\}=\{m_1\dots m_K\}$	 (14)
8	$\boldsymbol{A}_{\mathcal{M},f} = (\boldsymbol{a}_{m_1,f}\dots\boldsymbol{a}_{m_K,f})$ $\boldsymbol{A}_{\mathcal{M}} = \mathrm{diag}(\boldsymbol{A}_{\mathcal{M},1}\dots\boldsymbol{A}_{\mathcal{M},F})$	
9	$(\sigma^2)^{\mathrm{new}} = \frac{1}{N-K}\,\mathrm{tr}\left((\boldsymbol{I}_N - \boldsymbol{A}_{\boldsymbol{\mathcal{M}}}\boldsymbol{A}_{\boldsymbol{\mathcal{M}}}{}^+)\boldsymbol{S_y}\right)$	(28)
10	$\epsilon = \|\boldsymbol{\gamma}^{\mathrm{new}} - \boldsymbol{\gamma}^{\mathrm{old}}\|_1 / \|\boldsymbol{\gamma}^{\mathrm{old}}\|_1$	(30)
11	end	
12	Output: $\mathcal{M}$, $\boldsymbol{\gamma}^{\mathrm{new}}$, $(\sigma^2)^{\mathrm{new}}$	

Table 1. SBL Algorithm: In line 6 choose 23, 22 or 24.

2.5. 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 [14] provides an asymptotically efficient estimate of σ^2.

Let $\boldsymbol{\Gamma}_{\mathcal{M}} = \mathrm{diag}(\boldsymbol{\gamma}_{\mathcal{M}}^{\mathrm{new}})$ be the covariance matrix of the K active sources obtained above with corresponding active steering matrix $\boldsymbol{A}_{\mathcal{M}}$ which maximizes (16). The corresponding data covariance matrix is

$$\boldsymbol{\Sigma_y} = \sigma^2\boldsymbol{I}_{NF} + \boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H, \tag{25}$$

where $\boldsymbol{I}_N$ is the identity matrix of order N. The data covariance models (11) and (25) are identical. At the optimal solution $(\boldsymbol{\Gamma}_{\mathcal{M}}, \sigma^2)$, Jaffer's necessary condition ([15]:Eq.(6)) must be satisfied

$$\boldsymbol{A}_{\mathcal{M}}^H\left(\boldsymbol{S_y} - \boldsymbol{\Sigma_y}\right)\boldsymbol{A}_{\mathcal{M}} = \boldsymbol{0}. \tag{26}$$

Substituting (25) into (26) gives

$$\boldsymbol{A}_{\mathcal{M}}^H\left(\boldsymbol{S_y} - \sigma^2\boldsymbol{I}_{NF}\right)\boldsymbol{A}_{\mathcal{M}} = \boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}}. \tag{27}$$

Multiplying (27) from right and left with the pseudo inverse $\boldsymbol{A}_{\mathcal{M}}^+ = (\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{A}_{\mathcal{M}}^H$ and $\boldsymbol{A}_{\mathcal{M}}^{+H}$ respectively and subtracting $\boldsymbol{S_y}$ from both sides yields [14]

$$(\sigma^2)^{\mathrm{new}} = \frac{1}{N-K}\,\mathrm{tr}\left((\boldsymbol{I}_{NF} - \boldsymbol{A}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^+)\boldsymbol{S_y}\right). \tag{28}$$

This estimate requires $K < N$ and will underestimate the noise for small L.

Several estimators for the noise σ^2 are proposed based on EM [8, 11, 20, 12]. Empirically, neither of these converge

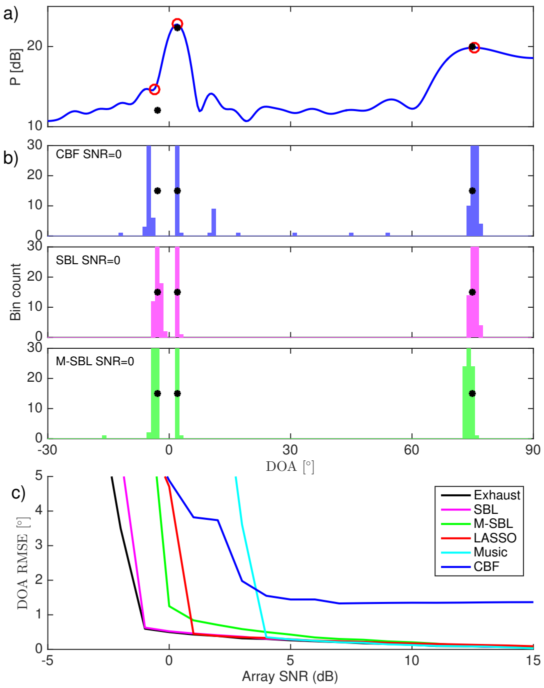

Fig. 1. Multiple L=50 snapshot example for sources at DOAs $[-3,\ 2,\ 75]^\circ$ with magnitudes $[12,\ 22,\ 20]$ dB. a) spectra for CBF and SBL (o) at SNR=0 dB. b) CBF, 23, and 24 histogram based on 100 Monte Carlo simulations at SNR=0 dB. c) RMSE performance versus array SNR for exhaustive, 23, 24, LASSO, MUSIC, and CBF. The true source positions (•) are indicated in a) and b).

well in our application. For a comparative illustration in Sec. 3 we use the iterative noise σ^2 EM estimate in [12],

$$(\sigma^2)^{\text{new}} = \frac{\frac{1}{L}\|\boldsymbol{Y}-\boldsymbol{A}\boldsymbol{\mu}_{\boldsymbol{X}}\|_{\mathcal{F}}^2 + (\sigma^2)^{\text{old}}(M-\sum_{i=1}^{M}\frac{(\Sigma_x)_{ii}}{\gamma_i})}{NF} \,. \quad (29)$$

2.6. SBL Algorithm

Given the observed $\boldsymbol{Y}$, we iteratively update $\boldsymbol{\mu}_{\boldsymbol{X}}$ (9) and $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ (11) by using the current $\boldsymbol{\gamma}$. Either 23, 22, or 24 can update γ_m for $m = 1, \ldots, M$ and then (28) is used to estimate σ^2. The algorithm is summarized in Table 1.

The convergence rate ϵ measures the relative improvement of the estimated total source power,

$$\epsilon = \|\boldsymbol{\gamma}^{\text{new}} - \boldsymbol{\gamma}^{\text{old}}\|_1 \Big/ \|\boldsymbol{\gamma}^{\text{old}}\|_1 \,. \quad (30)$$

The algorithm stops when $\epsilon \le \epsilon_{\min}$ and the output is the active set $\mathcal{M}$ (14) from which all relevant source parameter estimates are computed.

3. EXAMPLE

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 3 sources at DOAs $[-3,\ 2,\ 75]^\circ$ with magnitudes $[12,\ 22,\ 20]$ dB[9].

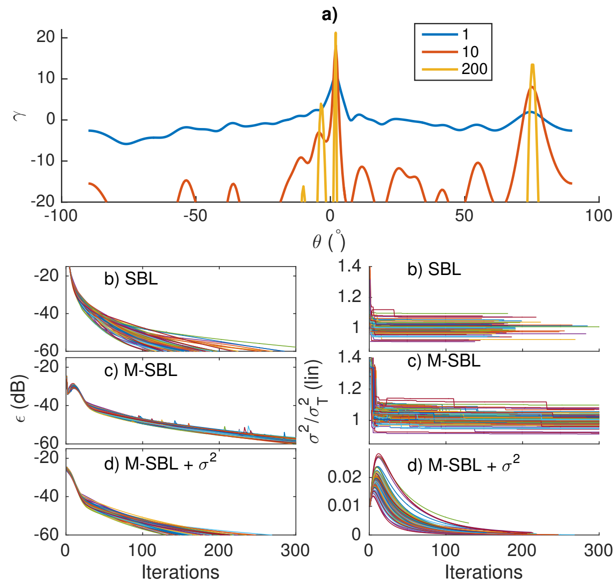

Fig. 2. Convergence at SNR=0 dB with L=50. a) $\boldsymbol{\gamma}$ at iteration 1, 10 , 200 for 23. Convergence of (b) 23 and (c and d) 24 for 100 Monte Carlo simulations. Convergence is shown for ϵ (left) and σ^2/σ_T^2 (right). In (b–c) the noise estimate $(\sigma^2)^{\text{new}}$ is based on (28) and in d) (29).

We consider an array with N=20 elements and half wavelength intersensor spacing. The DOAs are assumed to be on an angular grid $[-90{:}0.5{:}90]^\circ$, M=361, and L=50 snapshots are observed. The noise is modeled as iid complex Gaussian, though robustness to array imperfections [21] and extreme noise distributions [22] 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

$$\sigma_T^2 = \mathsf{E}[\|\boldsymbol{N}\|_{\mathcal{F}}^2]/L/N = 10^{-\text{SNR}/10}\ \mathsf{E}\,\frac{\|\boldsymbol{A}\boldsymbol{X}\|_{\mathcal{F}}^2}{LNF}. \quad (31)$$

The estimated $(\sigma^2)^{\text{new}}$ (28) deviates from σ_T^2 (31) randomly.

Figure 1 compares DOA estimation methods for the simulation. The LASSO solution is found considering multiple snapshots [9] and programmed in CVX[23]. 23 and 24 are calculated using the pseudocode on Table 1. CBF suffers from low-resolution and the effect of sidelobes in contrast to sparsity based methods as shown in the power spectra in Fig. 1a.

At array SNR=0 dB the histogram in Fig. 1b shows that CBF poorly locates the neighboring DOAs at broadside. 23 and 24 localize the sources well. The root mean squared error (RMSE) in Fig. 1c shows that CBF has low resolution as the main lobe is too broad (see Fig. 1a) and MUSIC per-

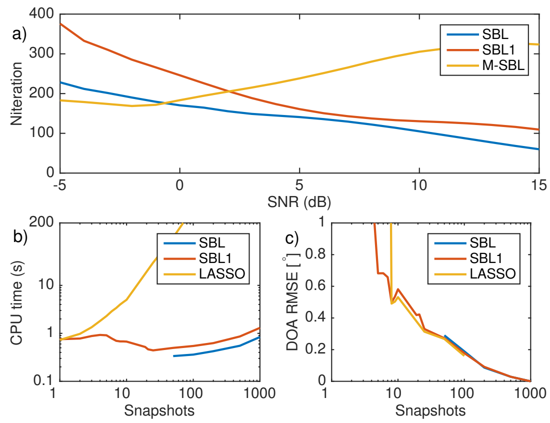

Fig. 3. a) Average number of iterations at each SNR for 24, 22, and 23 with L=50 snapshots. At array SNR=5 dB, b) average CPU time and c) RMSE for LASSO and 23 vs. number of snapshots. All results are an average of 100 Monte Carlo simulations.

forms well for $\mathrm{SNR}{>}5\,\mathrm{dB}$. For this case we include exhaustive search, which defines a lower performance bound and requires $361!/(3!358!){=}7.8{\cdot}10^6$ evaluations. LASSO and the SBL methods perform better than MUSIC and offer similar accuracy to the exhaustive search.

We compare the convergence of 23 and 24 at array SNR=0 dB (Fig. 2). The spatial spectrum (Fig. 2a) shows how the estimate $\boldsymbol{\gamma}$ improves with 23 iterations from initially locating only the main peak to locating also the weaker sources. 23 exhibits faster convergence than 24 to $\epsilon_{\min} = -60$ dB where the algorithm stops (Figs. 2b versus 2c). 24 underestimates σ^2 significantly when using (29) (Fig. 2d).

The average number of iterations for 23 and 22 decreases with SNR but increases for 24 (Fig. 3a). For 23 and 22 the CPU time (Macbook Pro 2014) is nearly constant with number of snapshots (Fig. 3b). The number of estimated parameters $(\boldsymbol{\gamma}, \sigma^2)$ is independent on the number of snapshots, but increasing the number of snapshots improves the estimation accuracy (lower RMSE). Contrarily, for LASSO the number of degrees of freedom in $\boldsymbol{X}$ increases as do CPU time with number of snapshots increases.

4. CONCLUSIONS

A sparse Bayesian learning (SBL) algorithm is derived for high-resolution DOA estimation from multi-snapshot complex-valued array data. Evidence maximization based on derivatives is used to estimate the source powers and the noise variance. The estimated source power at each potential DOA is used 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 while the computational effort is nearly independent of snapshots.

5. 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.*, **53**(8):3010–3022, 2005.

[2] G. F. Edelmann and C. F. Gaumond. Beamforming using compressive sensing. *J. Acoust. Soc. Am.*, **130**(4):232–237, 2011.

[3] C. F. Mecklenbräuker, P. Gerstoft, A. Panahi, and M. Viberg. Sequential Bayesian sparse signal reconstruction using array data. *IEEE Trans. Signal Process.*, **61**(24):6344–6354, 2013.

[4] S. Fortunati, R. Grasso, F. Gini, M. S. Greco, and K. LePage. Single-snapshot DOA estimation by using compressed sensing. *EURASIP J. Adv. Signal Process.*, **120**(1):1–17, 2014.

[5] A. Xenaki, P. Gerstoft, and K. Mosegaard. Compressive beamforming. *J. Acoust. Soc. Am.*, **136**(1):260–271, 2014.

[6] A. Xenaki and P. Gerstoft. Grid-free compressive beamforming. *J. Acoust. Soc. Am.*, **137**:1923–1935, 2015.

[7] H.L. Van Trees. *Optimum Array Processing*, chapter 1–10. Wiley-Interscience, New York, 2002.

[8] D. P. Wipf and B.D. Rao. An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Proc.*, **55**(7):3704–3716, 2007.

[9] P. Gerstoft, A. Xenaki, and C.F. Mecklenbräuker. Multiple and single snapshot compressive beamforming. *J. Acoust. Soc. Am.*, 138(4):2003–2014, 2015.

[10] P. Stoica and P. Babu. SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation. *Signal Proc.*, 92(7):1580–1590, 2012.

[11] D. P. Wipf and B.D. Rao. Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Proc*, 52(8):2153–2164, 2004.

[12] Z Zhang, TP Jung, S. Makeig, Z Pi, and BD. Rao. Spatiotemporal sparse Bayesian learning with applications to compressed sensing of multichannel physiological signals. *IEEE Trans Neural Syst Rehab Eng*, 22(6):1186–1197, 2014.

[13] JA. Zhang, Z. Chen, P. Cheng, and X. Huang. Multiple-measurement vector based implementation for single-measurement vector sparse Bayesian learning with reduced complexity. *Signal Proc.*, 118:153–158, 2016.

[14] J.F. Böhme. Source-parameter estimation by approximate maximum likelihood and nonlinear regression. *IEEE J. Oc. Eng.*, 10(3):206–212, 1985.

[15] AG. Jaffer. Maximum likelihood direction finding of stochastic sources: A separable solution. In *IEEE Int. Conf. Acoust., Speech, and Sig. Proc.*, volume 5, pages 2893–2896, 1988.

[16] P. Stoica and A. Nehorai. On the concentrated stochastic likelihood function in array processing. *Circuits Syst. Signal Proc.*, 14(5):669–674, 1995.

[17] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, **58**(1):267–288, 1996.

[18] J.F. Böhme. Estimation of spectral parameters of correlated signals in wavefields. *Signal Processing*, 11:329–337, 1986.

[19] AP Dempster, NM Laird, and DB Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Statist Soc B*, 39:1–38, 1977.

[20] M. E. Tipping. Sparse Bayesian learning and the relevance vector machine. *J. Machine Learning Research*, 1:211–244, 2001.

[21] C. Weiss and A.M. Zoubir. Doa estimation in the presence of array imperfections: A sparse regularization parameter selection problem. In *IEEE Workshop on Statis. Signal Proc.*, pages 348–351, June 2014.

[22] E. Ollila. Multichannel sparse recovery of complex-valued signals using Huber's criterion. In 3rd *Int. COSERA Workshop*, Pisa, Italy, 2015.

[23] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. cvxr.com/cvx, Last viewed 9 Feb 2016.