Signal Processing 161 (2019) 63–73

Contents lists available at ScienceDirect

Signal Processing

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

DOA Estimation in heteroscedastic noise

Peter Gerstoft[a,*], Santosh Nannuru[b], Christoph F. Mecklenbräuker[c], Geert Leus[d]

[a] University of California San Diego, La Jolla, CA 92093-0238, USA
[b] Signal Processing and Communications Research Center, IIIT, Hyderabad, India
[c] Institute of Telecommunications, TU Wien, Vienna 1040, Austria
[d] Dept. of Electrical Eng., Math. and Comp. Science, Delft Univ. of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 21 September 2018
Revised 6 January 2019
Accepted 16 March 2019
Available online 19 March 2019

Keywords:
Heteroscedastic noise
Sparse reconstruction
Array processing
DOA Estimation
Compressive beamforming
Phase-only processing

ABSTRACT

The paper considers direction of arrival (DOA) estimation from long-term observations in a very noisy environment. The concern is to derive methods obtaining reasonable DOAs at very low SNR. The noise is assumed zero-mean Gaussian and its variance varies in time and space, causing stationary data models to fit poorly over long observation times. Therefore a heteroscedastic Gaussian noise model is introduced where the variance varies across observations and sensors. The source amplitudes are assumed independent zero-mean complex Gaussian distributed with unknown variances (i.e. the source powers), inspiring stochastic maximum likelihood (ML) DOA estimation. The DOAs of plane waves are estimated from multi-snapshot sensor array data using sparse Bayesian learning (SBL) where the noise is estimated across both sensors and snapshots. This SBL approach is more flexible and performs better than other high-resolution methods since they cannot estimate the heteroscedastic noise process. An alternative to SBL is simple data normalization, whereby only the phase across the array is utilized. Simulations in noisy environments demonstrate that taking the heteroscedastic noise into account causes the DOA estimation to fail at lower SNR, often at 20 dB lower SNR.

1. Introduction

With long observation times, weak signals can be extracted in a noisy environment. Most statistical signal processing tools assume the noise to be wide sense stationary: the noise variance is constant. However, for long observation times the noise process may show non-stationary behavior, by the noise variance changing in both space and time. Such noise process is called heteroscedastic, meaning that the noise variance is changing. While the noise variance is a nuisance parameter that we are not interested in, it may still need to be estimated in the processing chain in order to obtain accurate estimates of the weak signals. We focus the paper in deriving and demonstrating methods where DOA estimation fails at much lower SNR then for a stationary noise assumption.

Accounting for the noise variation is important for machine learning [1,2] and related to robust statistics [3–5]. Heteroscedastic noise models have been used in e.g. finance [6] and image processing [7]. In statistical signal processing, the noise has been assumed to vary spatially [8–12], contain structure [13,14], or have outliers [15]. Inspired by [6], the heteroscedastic model was used to predict the time evolution of the noise for DOA estimation [16,17]. The proposed processing could be applied to spatial coherence loss [18–20] or to wavefront decorrelation, where turbulence causes the wave front to be incoherent for certain observations (thus more noisy). This has lead to so-called lucky imaging in astronomy [21] or lucky ranging in ocean acoustics [22], where only the measurements giving good results are used. As a result an involved hypothesis testing is needed to determine the measurements to be used. In contrast, we propose using all measurements.

The papers above [8–15,22] are mainly concerned with high SNR cases. In contrast, we are interested in the low SNR case where a more complete modelling of the noise nuisance parameter can cause the DOA estimation to fail $\sim$20dB later.

In applications, a simple way to account for noise power variations is to normalize the data to constant magnitude. The normalized data only contains phase information as demonstrated for beamforming in seismology [23,24], noise cross correlation in seismology [25–29] and acoustics [30], source deconvolution in ocean acoustics [31,32], and speaker localization [33,34]. An example of non-uniform spatiotemporal noise is a seismic array for finding weak sources where interference from a passing car dominates a

* Corresponding author.
E-mail addresses: gerstoft@ucsd.edu, pgerstoft@ucsd.edu (P. Gerstoft), cfm@ieee.org (C.F. Mecklenbräuker), g.j.t.leus@tudelft.nl (G. Leus).
URL: http://noiselab.ucsd.edu (P. Gerstoft)

https://doi.org/10.1016/j.sigpro.2019.03.014
0165-1684/

few nearby sensors [35] or a strong earthquake can dominate a few snapshots. Similarly, in ocean acoustics, a passing ocean wave on the surface temporarily increases the noise on the hydrophones near to the surface.

High-resolution DOA estimators such as MUSIC [36] which are based on the sample covariance matrix (SCM) do not perform well for heteroscedastic noise as few sensors or samples with loud noise realizations dominate the SCM. More robust methods than an eigenvalue decomposition are needed to separate the signal and noise subspaces. We demonstrate that for well-separated sources normalizing the data which amounts to retaining only the phase information works well.

When the sources are closely spaced, more advanced parametric methods are needed for DOA estimation when the noise power is varying in space and time and the sources are weak. In this paper we resolve closely spaced weak sources when the noise power is varying in space and time. Specifically, we derive noise variance estimates and demonstrate this for compressive beamforming [37–41] using multiple measurement vectors (MMV), also called multiple snapshots. We solve the MMV problem using the sparse Bayesian learning (SBL) framework [39,42,43]. We assume the source signals to jointly follow a zero-mean multivariate complex normal distribution with unknown power levels. The noise across sensors and snapshots also follows a zero-mean multivariate complex normal distribution with unknown variances. These assumptions lead to a Gaussian likelihood function. The corresponding posterior distribution is also Gaussian and already developed SBL approaches solve this well.

We base our development on our fast SBL method [42,43] which simultaneously estimates noise variances as well as source powers. For the heteroscedastic noise considered here, there could potentially be as many unknown variances as number of observations. Existing techniques are based on minimization-majorization [44] and expectation-maximization (EM) [39,40,45–49], though not all estimates work well. Instead, we estimate the unknown variances using approximate stochastic ML [50–52] modified to obtain noise estimates even for a single observation.

At high SNR, the Cramér-Rao bound (CRB) for spatially colored noise has been well-analysed [9]. However, our proposed methods are designed for DOA estimation from array data at low SNR and unknown noise variances which may vary in space and time. Due to the low SNR, the CRB does not provide a useful bound as we are in the threshold or no-information region where the CRB is not tight [53]. Without further assumptions on the noise variance evolution in space and time, the number of unknown noise variance parameters grows linearly with observation time and number of array elements. Thus, the ratio of the number of unknowns to the number of observations does not converge to zero when the observation time is increased. For this reason, we cannot expect that the CRB is attainable for large observation time or large array size even at high SNR.

To summarize, we develop methods for extracting information from weak stationary signals buried in non-stationary noise using long-term observations. We focus on DOA estimation for spatiotemporal heteroscedastic noise, the noise variance estimates are in Sections 3.1–3.3. We base our DOA analysis on SBL developed for stationary noise (Section 3.4) and augment it with heteroscedastic noise. This model can have more noise and signal parameters to estimate than observations, as there as many noise estimates as observations. The examples demonstrate that SBL with heteroscedastic noise works quite well (Section 5). Interestingly, empirical whitening (Section 4), as used by practitioners [23]–[34] works quite well at very low SNR for well-separated sources (Section 5).

2. Background

We treat the parameters of heteroscedastic noise as unknown deterministic variables as described in Section 2.1. These parameters are estimated as described in Sections 3.1—3.3. This is in contrast to modeling the noise parameters as random variables with an associated prior distribution. Such alternative model with random parameters could be treated by a hierarchical Bayesian approach, but we do not follow up on this here.

2.1. Heteroscedastic noise observation model

For the lth observation snapshot, we assume the linear model

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

where the dictionary $\mathbf{A} \in \mathbb{C}^{N\times M}$ is constant and known, and the source vector $\boldsymbol{x}_l \in \mathbb{C}^M$ contains the physical information of interest. Further, $\boldsymbol{n}_l \in \mathbb{C}^N$ is additive zero-mean circularly symmetric complex Gaussian noise, which is generated from a *heteroscedastic* Gaussian process $\boldsymbol{n}_l \sim \mathcal{CN}(\boldsymbol{n}_l; \mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{n}_l})$, where $\boldsymbol{\Sigma}_{\boldsymbol{n}_l}$ is the covariance matrix corresponding to the lth observation of the noise vector $\boldsymbol{n}_l$. Due to the circular symmetry of the noise the phase is uniformly distributed.

We specialize to diagonal covariance matrices, parameterized as

$$\boldsymbol{\Sigma}_{\boldsymbol{n}_l} = \sum_{n=1}^{N} \sigma_{nl}^2 \mathbf{J}_n = \mathrm{diag}(\sigma_{1l}^2, \ldots, \sigma_{Nl}^2), \tag{2}$$

where $\mathbf{J}_n = \mathrm{diag}(\boldsymbol{e}_n) = \boldsymbol{e}_n \boldsymbol{e}_n^T$ with $\boldsymbol{e}_n$ the nth standard basis vector. Note that the covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{n}_l}$ are varying over the snapshot index $l = 1, \ldots, L$ and we introduce the matrix of all noise standard deviations as

$$\mathbf{V_N} = \begin{pmatrix} \sigma_{11} & \ldots & \sigma_{1L} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \ldots & \sigma_{NL} \end{pmatrix} \in \mathbb{R}_{0+}^{N\times L}, \tag{3}$$

where $\mathbb{R}_{0+}$ denotes non-negative real numbers.

We consider three noise cases for the a priori knowledge on the noise covariance model (2) which are expressed as constraints on $\mathbf{V_N}$.

Case I We assume wide-sense stationarity of the noise in space and time: $\sigma_{nl}^2 = \sigma^2$. The model is homoscedastic,

$$\mathbf{V_N} \in \mathcal{V}_\mathrm{I} = \{\mathbf{V} \in \mathbb{R}_{0+}^{N\times L} | \mathbf{V} = \sigma \mathbf{1}_N \mathbf{1}_L^T\} \quad , \tag{4}$$

where $\mathbf{1}_N$ is the N-dimensional all-one vector.

Case II We assume wide-sense stationarity of the noise in space only, i.e., the noise variance for all sensor elements is equal across the array, $\sigma_{nl}^2 = \sigma_{0l}^2$ and it varies over snapshots. The noise variance is heteroscedastic in time (across snapshots),

$$\mathbf{V_N} \in \mathcal{V}_\mathrm{II} = \{\mathbf{V} \in \mathbb{R}_{0+}^{N\times L} | \mathbf{V} = \mathbf{1}_N [\sigma_{01} \ldots \sigma_{0L}]^T\}. \tag{5}$$

For Noise Case II, an example of a noise standard deviation matrix $\mathbf{V_N}$ is illustrated in Fig. 1a.

Case III No additional constraints other than (3). The noise variance is heteroscedastic across both time and space (sensors and snapshots). In this case

$$\mathbf{V_N} \in \mathcal{V}_\mathrm{III} = \mathbb{R}_{0+}^{N\times L}. \tag{6}$$

For Noise Case III, an example of a random $\mathbf{V_N}$ is illustrated in Fig. 1b and an example for a spatiotemporal evolution of the noise standard deviation is shown in Fig. 1c.

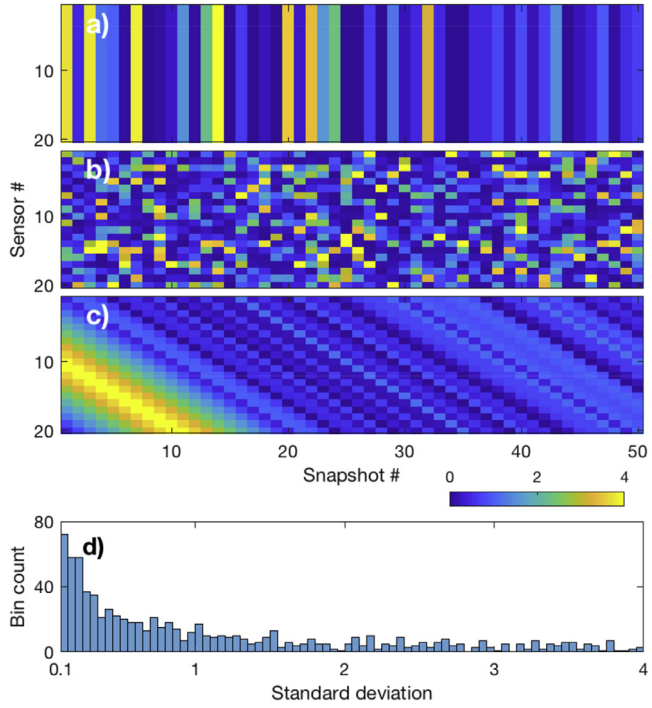

Fig. 1. Examples of noise standard deviation matrix $\mathbf{V_N}$ in (3) for $N = 20$ sensors and $L = 50$ snapshots. The average noise standard deviation is normalized to 1. Noise Cases a) II, b) III, c) III with evolving noise, d) Histogram of noise standard deviation for Noise Case III (1000 samples) in b) with $\log_{10}\sigma_{nl} \sim \mathcal{U}(-1, 1)$.

The relation between the domains for the three noise cases is $\mathcal{V}_{\rm I} \subset \mathcal{V}_{\rm II} \subset \mathcal{V}_{\rm III} = \mathbb{R}_{0+}^{N\times L}$. From the elements in the sets $\mathcal{V}_d$ (d is I, II, or III) both $\mathbf{V_N}$ in (3) and $\mathbf{\Sigma}_{\boldsymbol{n}_l}$ in (2) can be constructed.

2.2. Array model

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

$$\mathbf{Y} = \mathbf{AX} + \mathbf{N}. \tag{7}$$

The dictionary $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M] \in \mathbb{C}^{N\times M}$ contains the array steering vectors for all hypothetical DOAs as columns, with the (n, m)th element given by $\mathrm{e}^{\mathrm{i}\frac{\omega d_n}{c}\sin\theta_m}$ (d_n is the distance to the reference element and c the phase speed). The array configuration d_n is arbitrary.

We assume $M > N$ and thus (7) is underdetermined. In the presence of only few stationary sources, the source vector $\boldsymbol{x}_l$ is K-sparse with $K \ll M$. We define the lth active set

$$\mathcal{M}_l = \{m \in \mathbb{N} | x_{ml} \neq 0\}, \tag{8}$$

and assume $\mathcal{M}_l = \mathcal{M} = \{m_1, \ldots, m_K\}$ is constant across all snapshots l. Also, we define $\mathbf{A}_{\mathcal{M}} \in \mathbb{C}^{N\times K}$ which contains only the K "active" columns of $\mathbf{A}$. In the following, $\|\cdot\|_p$ 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 by examining the angular spectrum. The parameter K is required only for scaling the noise standard deviation in the SBL algorithm, with a factor $1/\sqrt{N-K}$, and for the size of the projection matrix. For real-world data sets, we have good experience with choosing K based on the observed number of DOAs. In this paper, K is assumed known.

2.3. Prior on the sources

We assume that the complex source amplitudes x_{ml} are independent both across snapshots and across DOAs and follow a zero-mean circularly symmetric complex Gaussian distribution with DOA-dependent variance γ_m, $m = 1, \ldots, M$,

$$p(x_{ml}; \gamma_m) = \begin{cases} \delta(x_{ml}), & \text{for } \gamma_m = 0 \\ \frac{1}{\pi\gamma_m}\mathrm{e}^{-|x_{ml}|^2/\gamma_m}, & \text{for } \gamma_m > 0 \end{cases}, \tag{9}$$

$$p(\mathbf{X}; \boldsymbol{\gamma}) = \prod_{l=1}^{L}\prod_{m=1}^{M} p(x_{ml}; \gamma_m) = \prod_{l=1}^{L} \mathcal{CN}(\boldsymbol{x}_l; \mathbf{0}, \mathbf{\Gamma}), \tag{10}$$

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

$$\mathbf{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}) = \mathsf{E}[\boldsymbol{x}_l \boldsymbol{x}_l^H; \boldsymbol{\gamma}], \tag{11}$$

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

$$\mathcal{M} = \{m \in \mathbb{N} | \gamma_m > 0\}. \tag{12}$$

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

2.4. Heteroscedastic noise estimate

For the low SNR cases considered, it is important to take the noise variance into account. We propose two methods for doing this:

- We estimate the spatiotemporal noise parameters (matrix $\mathbf{V_N}$) and the DOAs (active set $\mathcal{M}$) and power levels ($\boldsymbol{\gamma}$) using Stochastic Maximum Likelihood, see Section 3.
- We weight the array data with matrix $\mathbf{W}_l$ which is a square root of the estimated inverse noise covariance matrix (39) or suitable approximations there of for the case of low SNR, see Section 4.

3. Stochastic maximum likelihood

We here derive the well-known stochastic likelihood function [44,54–56]. Given the linear model (7) with Gaussian source (10) and noise (2) the array data $\mathbf{Y}$ is Gaussian with for each snapshot l the covariance $\mathbf{\Sigma}_{\boldsymbol{y}_l}$ given by

$$\mathbf{\Sigma}_{\boldsymbol{y}_l} = \mathsf{E}[\boldsymbol{y}_l \boldsymbol{y}_l^H] = \mathbf{\Sigma}_{\mathbf{n}_l} + \mathbf{A\Gamma A}^H \tag{13}$$

The probability density function of $\mathbf{Y}$ is thus given by

$$p(\mathbf{Y}) = \prod_{l=1}^{L} \mathcal{CN}(\boldsymbol{y}_l; \mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{y}_l}) = \prod_{l=1}^{L} \frac{\mathrm{e}^{-\mathbf{y}_l^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \mathbf{y}_l}}{\pi^N \det \mathbf{\Sigma}_{\boldsymbol{y}_l}}. \tag{14}$$

The L-snapshot log-likelihood for estimating $\boldsymbol{\gamma}$ and $\mathbf{V_N}$ is

$$\log p(\mathbf{Y}; \boldsymbol{\gamma}, \mathbf{V_N}) \propto -\sum_{l=1}^{L} \left(\mathbf{y}_l^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \mathbf{y}_l + \log\det \mathbf{\Sigma}_{\boldsymbol{y}_l}\right). \tag{15}$$

This likelihood function is identical to the Type II likelihood function (evidence) in standard SBL [40,42,45] which is obtained by treating $\boldsymbol{\gamma}$ as a hyperparameter. The Type II likelihood is obtained by integrating the likelihood function over the complex source amplitudes, cf. (29) in [42]. The stochastic maximum likelihood approach is used here as it is more direct.

The parameter estimates $\hat{\boldsymbol{\gamma}}$ and $\widehat{\mathbf{V}}_{\mathbf{N}}$ are obtained by maximizing the likelihood, leading to

$$(\hat{\boldsymbol{\gamma}}, \widehat{\mathbf{V}}_{\mathbf{N}}) = \underset{\boldsymbol{\gamma}\geq 0,\, \mathbf{V}_{\mathbf{N}}\in\mathcal{V}_d}{\arg\max} \log p(\mathbf{Y}; \boldsymbol{\gamma}, \mathbf{V}_{\mathbf{N}}), \tag{16}$$

where $\mathcal{V}_d$ is the feasible set of noise variances $\mathbf{V}_{\mathbf{N}}$ in (3) corresponding to the noise cases (d = I, II, or III, see Section 2.1). If $\boldsymbol{\gamma}$ and $\mathbf{V}_{\mathbf{N}}$, or $\boldsymbol{\Sigma}_{\mathbf{n}_l}$, are known, then the posterior for each snapshot l is $p(\mathbf{x}_l|\mathbf{Y}) = \mathcal{CN}(\boldsymbol{x}_l; \hat{\mathbf{x}}_l^{\mathrm{MAP}}, \boldsymbol{\Sigma}_{\boldsymbol{x}_l})$ with the posterior mean $\hat{\mathbf{x}}_l^{\mathrm{MAP}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}_l}$ [42,57],

$$\hat{\mathbf{x}}_l^{\mathrm{MAP}} = \boldsymbol{\Gamma}\mathbf{A}^H\boldsymbol{\Sigma}_{\boldsymbol{y}_l}^{-1}\mathbf{y}_l, \tag{17}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{x}_l} = \left(\mathbf{A}^H\boldsymbol{\Sigma}_{\boldsymbol{n}_l}^{-1}\mathbf{A} + \boldsymbol{\Gamma}^{-1}\right)^{-1}. \tag{18}$$

The diagonal elements of $\boldsymbol{\Gamma}$, i.e., $\boldsymbol{\gamma}$, control the row-sparsity of $\hat{\mathbf{x}}_l^{\mathrm{MAP}}$ as for $\gamma_m = 0$ the corresponding mth element of $\hat{\mathbf{x}}_l^{\mathrm{MAP}}$ becomes 0 too.

While there has been more focus on estimating the DOAs or $\boldsymbol{\gamma}$, noise is an important part of the physical system and a correct estimate is needed for good convergence properties. For example, in SBL the noise variance controls the sharpness of the peaks in the $\boldsymbol{\gamma}$ spectrum, with higher noise levels giving broader peaks. Thus, as we optimize the DOAs we expect the noise levels to decrease and the $\boldsymbol{\gamma}$ spectrum to become sharper.

In the following, we estimate the noise variance parameters for the three noise cases described in Section 2.1 and introduce simplifying approximations suitable at low signal to noise ratios. In Sections 3.1–3.3, we assume the support of $\boldsymbol{\gamma}$ to be known. More specifically, we let $\boldsymbol{\Gamma}_{\mathcal{M}} = \mathrm{diag}(\boldsymbol{\gamma}_{\mathcal{M}}^{\mathrm{new}})$ be the covariance matrix of the K active sources obtained by maximizing (15). The corresponding active steering matrix is $\mathbf{A}_{\mathcal{M}}$.

3.1. Noise estimate, Case I

Under Noise Case I, where $\boldsymbol{\Sigma}_{\boldsymbol{n}_l} = \sigma^2\mathbf{I}_N$ with $\mathbf{I}_N$ the identity matrix of size N, stochastic ML [47,50,52] provides an asymptotically efficient estimate of σ^2 if the set of active DOAs $\mathcal{M}$ is known. The data covariance matrix is then given by

$$\boldsymbol{\Sigma}_{\boldsymbol{y}_l} = \sigma^2\mathbf{I}_N + \mathbf{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\mathbf{A}_{\mathcal{M}}^H. \tag{19}$$

Note that, the data covariance matrices (13) and (19) are identical. Let us then define the projection matrix onto the subspace spanned by the active steering vectors

$$\mathbf{P} = \mathbf{A}_{\mathcal{M}}\mathbf{A}_{\mathcal{M}}^{+} = \mathbf{A}_{\mathcal{M}}(\mathbf{A}_{\mathcal{M}}^H\mathbf{A}_{\mathcal{M}})^{-1}\mathbf{A}_{\mathcal{M}}^H = \mathbf{P}^H = \mathbf{P}^2. \tag{20}$$

and the sample covariance matrix

$$\boldsymbol{S}_{\boldsymbol{y}} = \frac{1}{L}\boldsymbol{Y}\boldsymbol{Y}^H. \tag{21}$$

We use the approximate ML noise variance estimate $\hat{\sigma}^2$ obtained by [50,52]

$$\sigma^2 = \frac{\mathrm{tr}[\boldsymbol{\Sigma}_{\boldsymbol{y}_l} - \mathbf{P}\mathbf{S}_{\boldsymbol{y}}]}{N-K} = \frac{\mathrm{tr}[\mathbf{S}_{\boldsymbol{y}} - \mathbf{P}\mathbf{S}_{\boldsymbol{y}}] + \epsilon}{N-K} \tag{22}$$

$$\approx \frac{\mathrm{tr}[(\mathbf{I}_N - \mathbf{P})\mathbf{S}_{\boldsymbol{y}}]}{N-K} = \hat{\sigma}^2. \tag{23}$$

Here, we introduce the power estimation error $\epsilon = \mathrm{tr}[\boldsymbol{\Sigma}_{\boldsymbol{y}_l} - \mathbf{S}_{\boldsymbol{y}}]$. The estimate (23) has been used for SBL implementations [42,47].

The above approximation motivates the noise power estimate for Noise Case I (23), which is *error-free* if $\mathrm{tr}[\boldsymbol{\Sigma}_{\boldsymbol{y}}] = \mathrm{tr}[\mathbf{S}_{\boldsymbol{y}}]$, *unbiased* because $\mathrm{E}[\epsilon] = 0$, *consistent* since also its variance tends to zero for $L\to\infty$ [58], and *asymptotically efficient* as it approaches the CRLB for $L\to\infty$ [59]. The Noise Case I estimate (23) is valid even for one snapshot.

Table 1
SBL Algorithm.

0	Initialize: $\boldsymbol{\gamma}^{\mathrm{new}} = \mathrm{diag}[\boldsymbol{A}^{\mathrm{H}}\boldsymbol{S}_{\boldsymbol{y}}\boldsymbol{A}]$,	
	$\boldsymbol{\Sigma}_{\boldsymbol{n}_l}^{\mathrm{new}}$ = Eq. (23), (25), (49), or (29)	
	with $\boldsymbol{P} = \mathbf{0}, K = 0$	
	$\epsilon_{\min} = 0.001, \epsilon = 2\epsilon_{\min}, j = 0, j_{\max} = 100$	
1	while $(\epsilon > \epsilon_{\min})$ and $(j < j_{\max})$	
2	$\boldsymbol{\gamma}^{\mathrm{old}} = \boldsymbol{\gamma}^{\mathrm{new}}$, $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}^{\mathrm{old}})$, $\boldsymbol{\Sigma}_{\boldsymbol{n}_l}^{\mathrm{old}} = \boldsymbol{\Sigma}_{\boldsymbol{n}_l}^{\mathrm{new}}$	
3	$\boldsymbol{\Sigma}_{\boldsymbol{y}_l} = \boldsymbol{\Sigma}_{\boldsymbol{n}_l}^{\mathrm{old}} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^H$	(13)
4	γ_m^{new} use (36)	
5	$\mathcal{M} = \{m \in \mathbb{N} \mid \mathrm{K\ largest\ peaks\ in}\ \boldsymbol{\gamma}\}$	
	$= \{m_1 \dots m_K\}$	(12)
6	$\mathbf{A}_{\mathcal{M}} = (a_{m_1}, \dots, a_{m_K})$, $\boldsymbol{P} = \mathbf{A}_{\mathcal{M}}\mathbf{A}_{\mathcal{M}}^{+}$	
7	$\boldsymbol{\Sigma}_{\boldsymbol{n}_l}^{\mathrm{new}}$ = choose from (23), (25), (49), or (29)	
8	$\epsilon = \|\boldsymbol{\gamma}^{\mathrm{new}} - \boldsymbol{\gamma}^{\mathrm{old}}\|_1 / \|\boldsymbol{\gamma}^{\mathrm{old}}\|_1, \quad j = j+1$	(38)
9	Output: $\mathcal{M}$, $\boldsymbol{\gamma}^{\mathrm{new}}$, $\boldsymbol{\Sigma}_{\boldsymbol{n}_l}^{\mathrm{new}}$	

For the purpose of noise covariance matrix estimation at very low SNR, when the signals of interest are weak, we may ignore all signal power in (23). We set $K = 0$ and $\mathbf{P} = \mathbf{0}$ in (23) and use the sensor data SCM as a low-cost estimate of the noise covariance matrix, i.e.,

$$\hat{\sigma}^2 = \frac{\|\mathbf{Y}\|_F^2}{NL} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}_l} = \frac{\|\mathbf{Y}\|_F^2}{NL}\mathbf{I}_N. \tag{24}$$

We do not use (24) in the SBL algorithm in Table 1, but it motivates the approximate pre-whitening in Section 4.

3.2. Noise estimate, Case II

For Noise Case II, where $\boldsymbol{\Sigma}_{\boldsymbol{n}_l} = \sigma_l^2\mathbf{I}_N$, we apply (23) for each snapshot l individually, leading to

$$\hat{\sigma}_l^2 = \frac{\mathrm{tr}[(\mathbf{I}_N - \mathbf{P})\boldsymbol{y}_l\boldsymbol{y}_l^H]}{N-K} = \frac{\|(\mathbf{I}_N - \mathbf{P})\boldsymbol{y}_l\|_2^2}{N-K}. \tag{25}$$

Similarly to (24), we may ignore all signal power in (25) for very low SNR. We set $K = 0$ and $\mathbf{P} = \mathbf{0}$ in (25) and use the sensor data SCM as a low-cost estimate of the noise covariance matrix, i.e.,

$$\hat{\sigma}_l^2 = \frac{\|\boldsymbol{y}_l\|_2^2}{N} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}_l} = \frac{\|\boldsymbol{y}_l\|_2^2}{N}\mathbf{I}_N. \tag{26}$$

We do not use (26) in the SBL algorithm in Table 1, but it motivates the approximate pre-whitening in Section 4.

3.3. Noise estimate, Case III

Let us start from the definition of the noise covariance

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{n}_l} &= \mathrm{E}\left[(\boldsymbol{y}_l - \mathbf{A}\boldsymbol{x}_l)(\boldsymbol{y}_l - \mathbf{A}\boldsymbol{x}_l)^H\right] \\ &= \mathrm{E}\left[(\boldsymbol{y}_l - \mathbf{A}_{\mathcal{M}}\boldsymbol{x}_{\mathcal{M},l})(\boldsymbol{y}_l - \mathbf{A}_{\mathcal{M}}\boldsymbol{x}_{\mathcal{M},l})^H\right].\end{aligned} \tag{27}$$

This motivates plugging-in the single-observation signal estimate $\widehat{\boldsymbol{x}}_{\mathcal{M},l} = \mathbf{A}_{\mathcal{M}}^{+}\boldsymbol{y}_l \in \mathbb{C}^K$ for the active (non-zero) entries in $\boldsymbol{x}_l$. This estimate is based on the single observation $\boldsymbol{y}_l$ and the projection matrix (20), giving the rank-1 estimate

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}_l} = (\mathbf{I} - \mathbf{P})\boldsymbol{y}_l\boldsymbol{y}_l^H(\mathbf{I} - \mathbf{P}). \tag{28}$$

Since the signal estimate $\widehat{\boldsymbol{x}}_{\mathcal{M},l}$ maximizes the estimated signal power, this noise covariance estimate is biased and the noise level is likely underestimated.

Since we assume independent noise across sensors, all off-diagonal elements of $\boldsymbol{\Sigma}_{\boldsymbol{n}_l}$ are known to be zero. With this constraint in mind, we modify (28) as

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}_l} &= \mathrm{diag}[\sigma_{1l}^2, \dots, \sigma_{nl}^2, \dots, \sigma_{Nl}^2] \\ &= \mathrm{diag}\left[\mathrm{diag}\left[(\mathbf{I} - \mathbf{P})\boldsymbol{y}_l\boldsymbol{y}_l^H(\mathbf{I} - \mathbf{P})\right]\right].\end{aligned} \tag{29}$$

The estimate (29) is demanding as for all the $N \times L$ complex-valued observations in $\mathbf{Y}$, we obtain $N \times L$ estimates of the noise variance.

Note that the estimate $\widehat{\mathbf{\Sigma}}_{\boldsymbol{n}_l}$ in (28) is not invertible whereas the diagonal constraint in (29) leads to a non-singular estimate of $\mathbf{\Sigma}_{\boldsymbol{n}_l}$ with high probability (it is singular only if an element of $\boldsymbol{y}_l$ is 0). As a result, the expression (29) for $\mathbf{\Sigma}_{\boldsymbol{y}_l}$ that is used for estimating $\boldsymbol{\gamma}$ in (36) is likely invertible.

Similarly to (24) and (26), we may ignore all signal power in (29) for very low SNR. We set $\mathbf{P} = \mathbf{0}$ in (29), resulting in the low-cost estimate

$$\widehat{\mathbf{\Sigma}}_{\boldsymbol{n}_l} = \mathrm{diag}\left[\mathrm{diag}\left[\boldsymbol{y}_l \boldsymbol{y}_l^H\right]\right] = \mathrm{diag}(|y_{1,l}|^2, \ldots, |y_{N,l}|^2). \tag{30}$$

or

$$\widehat{\sigma}_{nl}^2 = |y_{nl}|^2. \tag{31}$$

This can be shown to be the stochastic ML noise estimate for no sources ($K = 0$) or very low power sources. We do not use (30) in the SBL algorithm in Table 1, but it motivates the approximate pre-whitening in Section 4.

3.4. Source power estimation

We impose the diagonal structure $\mathbf{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$, in agreement with (10), and form derivatives of (15) with respect to the diagonal elements γ_m, cf [54]. Using

$$\frac{\partial \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1}}{\partial \gamma_m} = -\mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \frac{\partial \mathbf{\Sigma}_{\boldsymbol{y}_l}}{\partial \gamma_m} \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} = -\mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m \boldsymbol{a}_m^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1}, \tag{32}$$

$$\frac{\partial \log\det(\mathbf{\Sigma}_{\boldsymbol{y}_l})}{\partial \gamma_m} = \mathrm{tr}\left(\mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \frac{\partial \mathbf{\Sigma}_{\boldsymbol{y}_l}}{\partial \gamma_m}\right) = \boldsymbol{a}_m^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m, \tag{33}$$

the derivative of (15) is formulated as

$$\begin{aligned}
\frac{\partial \log p(\mathbf{Y}; \boldsymbol{\gamma}, \mathbf{V_N})}{\partial \gamma_m} &= \sum_{l=1}^{L} \left(\boldsymbol{a}_m^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{y}_l \boldsymbol{y}_l^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m - \boldsymbol{a}_m^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m\right) \\
&= \sum_{l=1}^{L} \boldsymbol{a}_m^H \left(\mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{y}_l \boldsymbol{y}_l^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} - \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1}\right) \boldsymbol{a}_m \\
&= \sum_{l=1}^{L} |\boldsymbol{y}_l^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m|^2 - \sum_{l=1}^{L} \boldsymbol{a}_m^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m \\
&= \frac{1}{\gamma_m^2} \sum_{l=1}^{L} |\hat{\mathbf{x}}_{ml}^{\mathrm{MAP}}|^2 - \sum_{l=1}^{L} \boldsymbol{a}_m^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m,
\end{aligned} \tag{34}$$

where the first term in the last equation is expressed by the first moment of the posterior $\hat{\mathbf{x}}_{ml}^{\mathrm{MAP}}$ (17) and γ_m for the mth element. Assuming $\hat{\mathbf{x}}_{ml}^{\mathrm{MAP}}$ and $\mathbf{\Sigma}_{\boldsymbol{y}_l}$ given (from previous iterations or initialization) and forcing (34) to zero gives the γ_m update:

$$(\gamma_m^{\mathrm{new}})^2 = \frac{\sum_{l=1}^{L} |\hat{\mathbf{x}}_{ml}^{\mathrm{MAP}}|^2}{\sum_{l=1}^{L} \boldsymbol{a}_m^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m}. \tag{35}$$

Since γ_m^{old} is common in $\hat{\mathbf{x}}_{ml}^{\mathrm{MAP}}$ across all L snapshots, see (17), we obtain the following

$$\gamma_m^{\mathrm{new}} = \gamma_m^{\mathrm{old}} \left(\frac{\sum_{l=1}^{L} |\boldsymbol{y}_l^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m|^2}{\sum_{l=1}^{L} \boldsymbol{a}_m^H \mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1} \boldsymbol{a}_m}\right)^{0.5}. \tag{36}$$

This update rule was used in [42]. Using an exponent of 1 (instead of 0.5) in (36) gives the update equation used in [39,43,57].

3.5. Sparse Bayesian learning

We are now ready to describe the SBL algorithm solving (16) without pre-whitening. For a detailed derivation of the source power estimates $\hat{\boldsymbol{\gamma}}$, see [42], and for multiple dictionaries see [43]. The algorithm iterates between the source power estimates $\hat{\boldsymbol{\gamma}}$ derived in Section 3.4 and the noise parameter estimates $\widehat{\mathbf{V}}_{\mathbf{N}}$ computed in Sections 3.1–3.3.

The proposed SBL algorithm is summarized in Table 1 which is applicable for Noise Cases I, II, and III. Given the observed $\mathbf{Y}$, we iteratively update $\mathbf{\Sigma}_{\boldsymbol{y}_l}$ (13) by using the current $\boldsymbol{\gamma}$ and $\mathbf{\Sigma}_{\boldsymbol{n}_l}$. The $\mathbf{\Sigma}_{\boldsymbol{y}_l}^{-1}$ is computed directly as the numerical inverse of $\mathbf{\Sigma}_{\boldsymbol{y}_l}$. For updating γ_m, $m = 1, \ldots, M$ we use (36). For the initialization of $\boldsymbol{\gamma}$ we use the conventional beamformer (CBF) power estimate

$$\boldsymbol{\gamma} = \mathrm{diag}[\boldsymbol{A}^{\mathrm{H}} \boldsymbol{S}_{\boldsymbol{y}} \boldsymbol{A}]. \tag{37}$$

The noise covariance matrix $\mathbf{\Sigma}_{\boldsymbol{n}_l}$ is based on either Noise Case I (23), II (25), or III (29). The noise is initialized using (23), (25), or (29) with $\boldsymbol{P} = \mathbf{0}, K = 0$, which provides an over estimate of the noise variance.

The convergence rate ϵ measures the relative change in estimated total source power,

$$\epsilon = \|\boldsymbol{\gamma}^{\mathrm{new}} - \boldsymbol{\gamma}^{\mathrm{old}}\|_1 \,/\, \|\boldsymbol{\gamma}^{\mathrm{old}}\|_1. \tag{38}$$

The algorithm stops when $\epsilon \le \epsilon_{\min}$ and the output is the active set $\mathcal{M}$ (see (12)) from which all source parameters are computed.

4. Approximate pre-whitening

The purpose of this section is to motivate the empirical evidence [23]–[34] that phase-only processing with conventional beamforming (CBF) might provide improved DOA estimates over using both amplitude and phase of the array data (not needed for SBL which estimates the heteroscedastic noise variance). Such improvements in DOA estimation are observed when sources are not closely spaced and at low SNR as demonstrated in the examples, Section 5.

For snapshot l, we factorize the inverse noise covariance $\mathbf{\Sigma}_{\boldsymbol{n}_l}^{-1} = \mathbf{W}_l^H \mathbf{W}_l$, where $\mathbf{W}_l$ is chosen to be square and positive definite. For heteroscedastic noise modeled in (2), we choose

$$\mathbf{W}_l = \mathrm{diag}(\sigma_{1l}^{-1}, \ldots, \sigma_{Nl}^{-1}). \tag{39}$$

The matrix $\mathbf{W}_l$ is useful for pre-whitening the sensor data. The corresponding whitened sensor data is

$$\widetilde{\boldsymbol{y}}_l = \mathbf{W}_l \boldsymbol{y}_l. \tag{40}$$

For known diagonal noise covariance $\mathbf{\Sigma}_{\boldsymbol{n}_l}$ the above means we have to normalize each row in (1) with σ_{nl} as then the noise satisfies $\widetilde{\boldsymbol{n}}_l \sim \mathcal{CN}(\widetilde{\boldsymbol{n}}_l; \mathbf{0}, \mathbf{I})$, and thus all entries are identically distributed.

For real-world sensor data, the noise covariance matrices are unknown and must be estimated as described in Sections 3.1–3.3. At very low SNR, we may use the corresponding low-cost estimates (24), (26), and (30). The corresponding pre-whitening matrices are then

$$\mathbf{W}_l \approx \begin{cases} \frac{\sqrt{NL}}{\|\mathbf{Y}\|_F} \mathbf{I} & \text{for Case I} \\ \frac{\sqrt{N}}{\|\boldsymbol{y}_l\|_2} \mathbf{I} & \text{for Case II} \\ \mathrm{diag}(|y_{1,l}|, \ldots, |y_{N,l}|)^{-1} & \text{for Case III} \end{cases} \tag{41}$$

leading to the approximately pre-whitened array data

$$\widetilde{\boldsymbol{y}}_l \approx \begin{cases} \frac{\sqrt{NL}}{\|\mathbf{Y}\|_F} \boldsymbol{y}_l & \text{for Case I} \\ \frac{\sqrt{N}}{\|\boldsymbol{y}_l\|_2} \boldsymbol{y}_l & \text{for Case II} \\ \left[y_{1,l}/|y_{1,l}|, \ldots, y_{N,l}/|y_{N,l}|\right]^T & \text{for Case III.} \end{cases} \tag{42}$$

Empirically, it has been found that applying pre-whitening to the data $\boldsymbol{y}_l$ only (the dictionary $\mathbf{A}$ is non-whitened) is effective in finding the strongest DOA [23]–[34].

For Noise Case I, pre-whitening does not play a role and the conventional beamformer is formulated using the power spectrum at DOA θ_m

$$P_{\mathrm{CBF}}(\theta_m) = \frac{1}{L} \boldsymbol{a}_m^H \boldsymbol{Y} \boldsymbol{Y}^H \boldsymbol{a}_m = \boldsymbol{a}_m^H \boldsymbol{S}_{\boldsymbol{y}} \boldsymbol{a}_m \tag{43}$$

where the sample covariance matrix $\boldsymbol{S_y}$ is defined in (21).

For Noise Cases II and III on the other hand, we will work with the approximately pre-whitened data. In those cases, the conventional beamformer leads to the powerspectra P_{CBF2} and P_{CBF3}, respectively, which can be formulated as

$$P_{\mathrm{CBF}^*}(\theta_m) = \frac{1}{L}\boldsymbol{a}_m^H \tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^H \boldsymbol{a}_m = \boldsymbol{a}_m^H \boldsymbol{S}_{\tilde{\boldsymbol{y}}} \boldsymbol{a}_m \tag{44}$$

where P_{CBF^*} is either P_{CBF2} or P_{CBF3}. The whitened sample covariance matrix is

$$\mathbf{S}_{\tilde{\boldsymbol{y}}} = \frac{1}{L}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^H = \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{y}_l \mathbf{W}_l^2 \boldsymbol{y}_l^H. \tag{45}$$

The weighting $\mathbf{W}_l$ is given by (41). For Noise Case III, only phase is used as can be observed from (42), which results in phase-only processing P_{CBF3}. As demonstrated in Section 5.2 this simple pre-whitening can improve the DOA performance at low SNR.

5. Examples

When the noise process is stationary (Noise Case I) and only a single source is active, the peak of the CBF spectrum provides the optimal DOA estimate. For heteroscedastic noise (Noise Cases II and III) this does not hold true in general. In the following simulation examples, we investigate different DOA estimators in heteroscedastic noise.

In Section 5.2, we investigate CBF (with and without weighting) and compare with MUSIC results. Next, we investigate a DOA estimator based on SBL, which takes into account the heteroscedastic noise explicitly (Section 5.3) for a single source and the case of multiple sources is discussed in Section 5.4.

5.1. Simulation setup

In the analysis of seismic data the noise for each snapshot was observed to be log-normal distributed [35]. Noise has also been modeled with extreme-value distributions [60]. In the simulations here, the noise follows a normal-uniform hierarchical model. The noise is complex zero-mean Gaussian with the noise standard deviation uniformly distributed over two decades, i.e., $\log_{10}\sigma_{nl} \sim \mathcal{U}(-1,1)$, where $\mathcal{U}$ is the uniform distribution. Three noise cases are simulated:

(a) Case I: constant noise standard deviation over snapshots and sensors,

(b) Case II: standard deviation changes across snapshots with $\log_{10}\sigma_l \sim \mathcal{U}(-1,1)$, and

(c) Case III: standard deviation changes across both snapshots and sensors with $\log_{10}\sigma_{nl} \sim \mathcal{U}(-1,1)$.

A realization of the noise standard deviation matrix is shown for Noise Case II in Fig. 1a. A realization for Noise Case III is shown in Fig. 1b) for $N = 20$ sensors and $L = 50$ snapshots. The corresponding histogram is presented in Fig. 1d. An example for a spatiotemporal evolution of the noise standard deviation is shown in Fig. 1c.

5.2. Single DOA with CBF and MUSIC

The single source is located at -3°. The complex source amplitude is stochastic and there is additive heteroscedastic Gaussian noise with an SNR variation from -40 to 0 dB. The sensor array has $N = 20$ elements with half wavelength spacing. We process $L = 50$ snapshots. The angle space $[-90, 90]^\circ$ is divided into a 0.5° grid ($M = 360$). The single-snapshot array signal-to-noise ratio (SNR) is

$$\mathrm{SNR} = 10\log_{10}[\mathrm{E}\left\{\|\mathbf{A}\boldsymbol{x}_l\|_2^2\right\}/\mathrm{E}\left\{\|\boldsymbol{n}_l\|_2^2\right\}]. \tag{46}$$

We first compute the beam spectra using CBF (43), as well as CBF2 and CBF3, both using (44) but with different weightings (42). When the noise is homoscedastic (constant standard deviation), the beam spectra for the three processors behave similarly (first row in Fig. 2). For heteroscedastic noise CBF2 and CBF3 give much better discrimination between signal and noise, see Fig. 2 middle (Noise Case II) and bottom row (Noise Case III).

The root mean squared error (RMSE) of the DOA estimates over $N_{\mathrm{sim}} = 500$ runs with random noise realizations is used for evaluating the algorithm

$$\mathrm{RMSE} = \sqrt{\sum_{n=1}^{N_{\mathrm{sim}}}\sum_{i=1}^{K}\frac{(\hat{\theta}_i^n - \theta_i^0)^2}{N_{\mathrm{sim}}K}}, \tag{47}$$

where θ_i^0 is the true DOA and $\hat{\theta}_i^n$ is the estimated DOA for the ith source and nth simulation when K sources are present.

The SNR curves (Fig. 3) demonstrate increased robustness of CBF2 and CBF3, failing 20 dB later. Due to the heteroscedastic noise, MUSIC performs worse than CBF for a single source. CBF3 diverges at an SNR 15–20 dB later than CBF for Noise Cases II and III.

5.3. Single DOA estimation with SBL

We use the following SBL methods with $\boldsymbol{\gamma}$ update (36) and parameters given in Table 1:

SBL: Solves Noise Case I using standard SBL, with σ from (23).
SBL2: Solves Noise Case II, with σ_l for each snapshot from (25).
SBL3: Solves Noise Case III, with σ_{nl} from (29).

Parameters used for SBL are in Table 1. In addition to these methods, we use basis pursuit (BP) as implemented in [41].

For Noise Case I, all the methods fail near the same SNR of -12.5 dB, Fig. 4a. When the noise is heteroscedastic across snapshots, CBF, BP, and SBL fail early. Since both SBL2 and SBL3 correctly model the noise, they perform the best for Noise Case II, Fig. 4b. CBF3 also performs well. For heteroscedastic Noise Case III, SBL3 fails the last and as before CBF3 also performs well. SBL3 fails 15 dB later than any other SBL-method. This demonstrates the usefulness of accurate noise models in DOA processing.

For Noise Cases II and III, the SCM (21) based on the raw array data $\mathbf{Y}$ might be dominated by a few ("corrupted") snapshots/sensors with strong noise. For CBF it is clear that a more robust approach is to dampen the influence of the corrupted snapshots/sensors by replacing the SCM (21) with the approximately pre-whitened SCM (45). This is the intuitive explanation that CBF3 performs well in Fig. 4. For MUSIC, an SCM (21) dominated by few corrupted snapshots/sensors will not be able to separate the signal and noise subspace in Fig. 4. Though basis pursuit (BP) is a compressive sensing method, due to changing noise, the regularization parameter cannot be uniquely chosen and hence its performance suffers in Fig. 4.

An example statistic of the heteroscedastic noise standard deviation is shown in Fig. 5 for a 20-element array with a single source. The standard deviation for each sensor is either 0 or $\sqrt{2}$. SBL3 estimates the standard deviation from (29). The average noise in Fig. 5b closely resembles the true noise (Fig. 5a) whereas the sample standard deviation estimate (Fig. 5c) has high variability since each estimate is based on just one observation. Given many simulations and snapshots, however, the mean of the estimated standard deviation is close to the true noise (Fig. 5d). On average, the noise estimate is close to the true noise.

5.4. Three DOA estimation with SBL

In Fig. 6, we consider three sources located at $[-3, 2, 50]^\circ$ with power [10, 22, 20] dB. For Fig. 6c, the CBF3 performs worse than

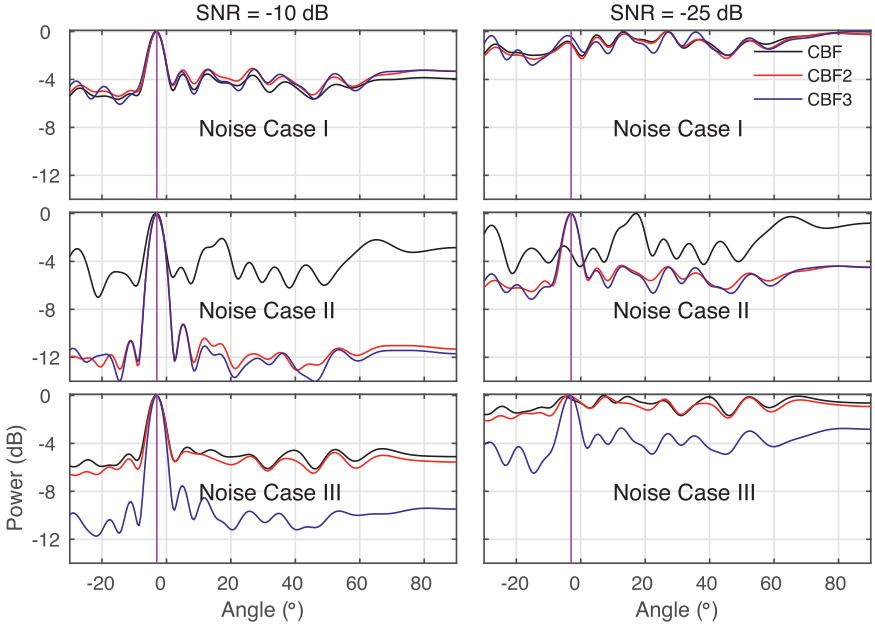

Fig. 2. Beam spectra for source at -3° for SNR -10 dB (left) and -25 dB (right) for CBF (black), CBF2 (red), CBF3 (blue). The noise standard deviation is a) constant or heteroscedastic with either b) $\log_{10}\sigma_l \sim \mathcal{U}(-1,1)$ or c) $\log_{10}\sigma_{nl} \sim \mathcal{U}(-1,1)$. 20 elements and 50 snapshots based on one simulation are used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

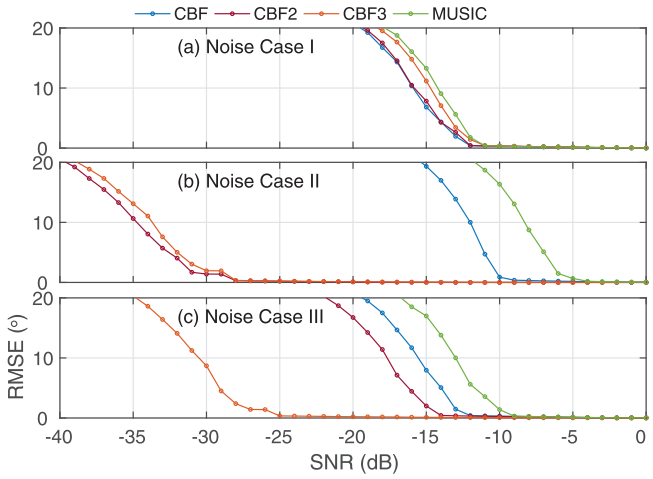

Fig. 3. Single source at $\theta = -3^\circ$: Array RMSE using beamforming with pre-whitening for Noise Cases (a) I, (b) II, and (c) III. Each noise case is solved with CBF, CBF2, CBF3, and MUSIC.

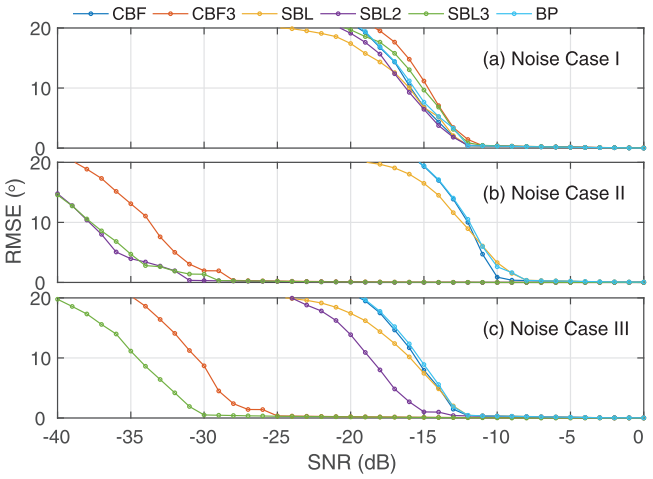

Fig. 4. Single source at -3°: RMSE vs. SNR for DOA estimation using various algorithms for Noise Cases (a) I, (b) II, and (c) III.

SBL3 from an SNR above -15 dB, as the sources at -3° and 2° are both in the CBF main lobe. Interestingly, CBF3 performs the best at very low SNR, below -15 dB where the RMSE is about 10°.

The localization ability of the algorithms can also be gauged from the histogram of the top three peaks, see Fig. 7. Since SBL3 accounts for the heteroscedastic Noise Case III, its histogram is concentrated near the true DOAs.

5.4.1. Noise estimate convergence

The performance of the SBL methods is strongly related to the accuracy of the noise estimates. For the lth snapshot, the true noise power $\sigma^2_{l,T}$ (Noise Case II) is

$$\sigma^2_{l,T} = \mathsf{E}[\|\boldsymbol{n}_l\|_2^2]/N = 10^{-\mathrm{SNR}/10}\ \mathsf{E}[\|\boldsymbol{A}\boldsymbol{x}_l\|_2^2]/N. \tag{48}$$

The estimated σ^2_l (25) deviates from $\sigma^2_{l,T}$ (48) randomly. Several alternative estimators for the noise variance are proposed based on EM [39,45,46,57,61]. We use the iterative noise EM-estimate from [45], given by

$$(\sigma_l^2)^{\mathrm{new}} = \frac{||\boldsymbol{y}_l - \boldsymbol{A}\widehat{\boldsymbol{x}}_l^{\mathrm{MAP}}||_F^2 + (\sigma_l^2)^{\mathrm{old}}\left(M - \sum_{i=1}^{M}\frac{(\boldsymbol{\Sigma}_{\boldsymbol{x}_l})_{ii}}{\gamma_i}\right)}{N}, \tag{49}$$

where the posterior mean $\widehat{\boldsymbol{x}}_l^{\mathrm{MAP}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}_l}$ are given in (17) and (18), respectively. Empirically, EM noise estimates such as (49) significantly underestimate the true noise variance in our applications.

Fig. 8 compares the two noise estimates for noise generated with $\log\sigma_l \sim \mathcal{U}(-1,1)$. The evolution of the histograms of the relative noise variance $\sigma^2_{\mathrm{Est}}/\sigma^2_T$ with iterations is in Fig. 8b for just 10 simulations, in total 10*50=500 variances are estimated. The mean of the ratio of $\sigma^2_{\mathrm{Est}}/\sigma^2_T$ is close to 1 for SBL2 but is much lower for the EM noise estimate (49), this likely causes the SBL2 (EM noise, (49)) to fail 3 dB earlier.

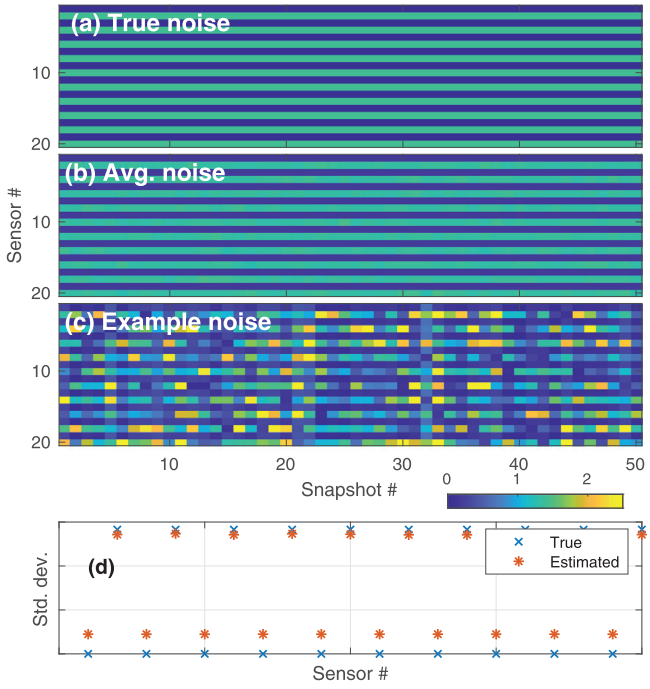

Fig. 5. Single source at -3°, array SNR=0 dB: SBL3 noise standard deviation matrix $\mathbf{V_N}$ in (3): (a) true standard deviations, (b) average estimated (500 simulations) standard deviations, (c) a typical estimate of standard deviations, and (d) average standard deviations across simulations and snapshots.

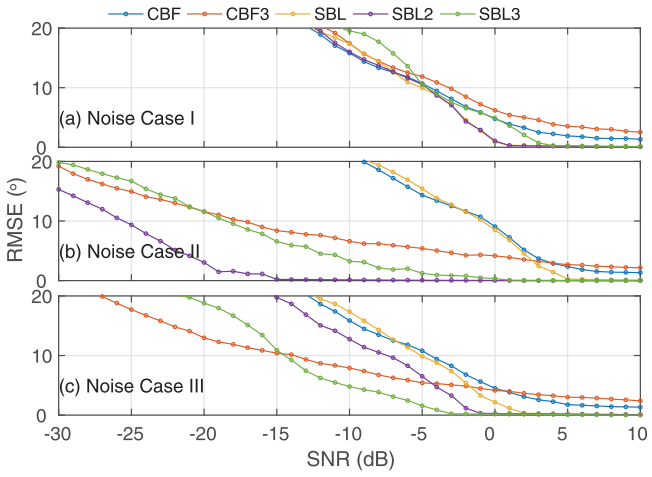

Fig. 6. RMSE vs. SNR with the three sources at $\{-3^\circ, 2^\circ, 50^\circ\}$ and power {10, 22, 20} dB.

5.4.2. Number of snapshots

The RMSE versus the number of snapshots for Noise Case III (Fig. 9) shows that SBL3 is most accurate with CBF3 following.

5.4.3. Noise distribution

For Noise Case III, we are just using one observation to estimate the noise standard deviation for SBL3 (29). Thus the individual estimates are not good as shown in the histogram in Fig. 10. The histograms are based on $N = 20$ sensors and $L = 50$ snapshots (1000 noise standard deviations). The true noise standard deviation histogram is generated from $\log_{10}\sigma_{T,nl} \sim \mathcal{U}(-1, 1)$, see Fig. 10a. The histogram of the deviation from the true standard deviation $\sigma_{\text{Est}} - \sigma_T$ (Fig. 10c) is well-centered (mean 0.007), but with large outliers.

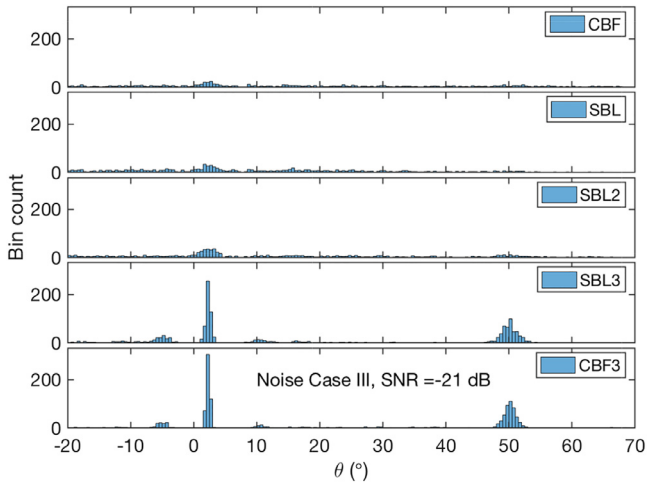

Fig. 7. Three sources at $\{-3^\circ, 2^\circ, 50^\circ\}$: Histogram of the top three peak locations for Noise Case III at SNR −15 dB.

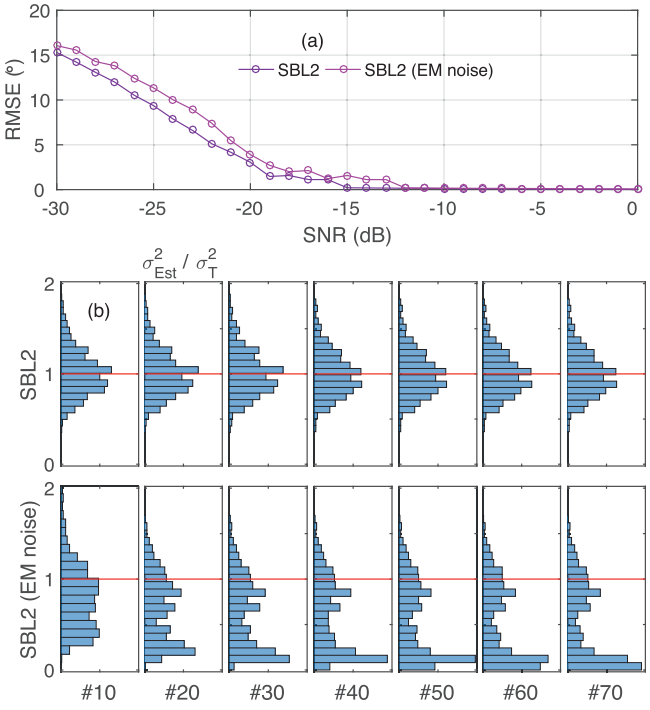

Fig. 8. Three sources at $\{-3^\circ, 2^\circ, 50^\circ\}$ (Noise Case II): (a) RMSE vs. SNR for SBL2 with two noise estimates. (b) Evolution of histogram of noise variance $\sigma^2_{\text{Est}}/\sigma^2_T$ for SBL2 (25) and SBL2 with EM noise estimate (49) with iterations at SNR −10 dB.

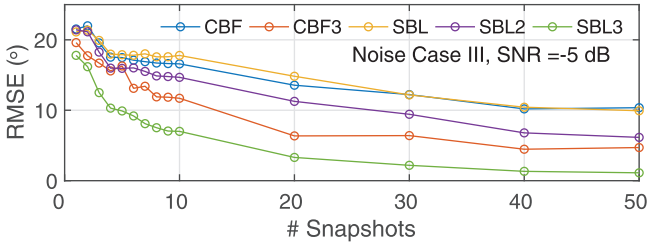

Fig. 9. Three sources at $\{-3^\circ, 2^\circ, 50^\circ\}$ (Noise Case III): RMSE vs. Number of snapshots with SNR −5 dB.

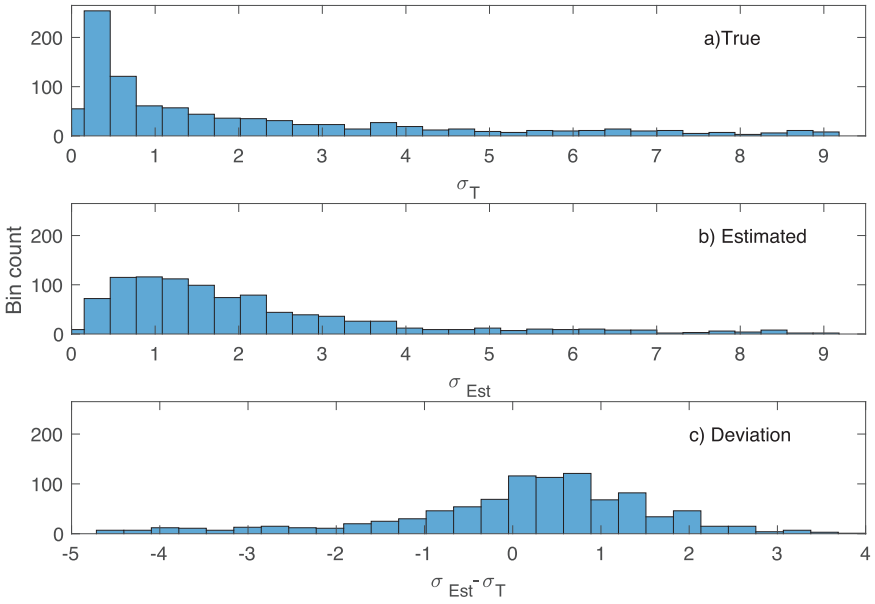

Fig. 10. Three sources at $\{-3^\circ, 2^\circ, 50^\circ\}$ (Noise Case III): Histogram of noise for a realization for SBL3 with 50 snapshots and 20 sensors (1000 observations) with SNR 0 dB. a) True σ_T, b) estimated σ_{Est}, and c) deviation $\sigma_{\text{Est}} - \sigma_T$.

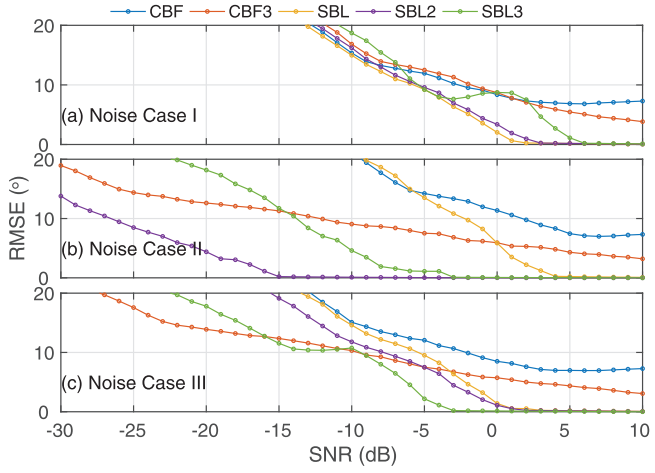

Fig. 11. As Fig. 6, but with the SBL implementation in [39].

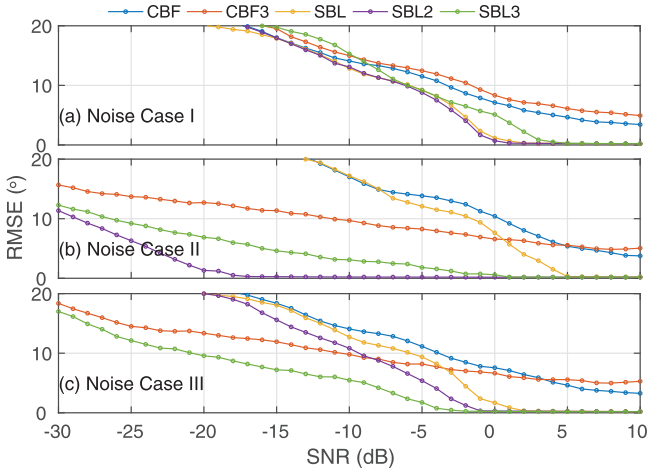

Fig. 12. RMSE vs. SNR with three off-grid sources at $\{-2.889^\circ, 2.222^\circ, 50.178^\circ\}$ and different power {10, 22, 20} dB.

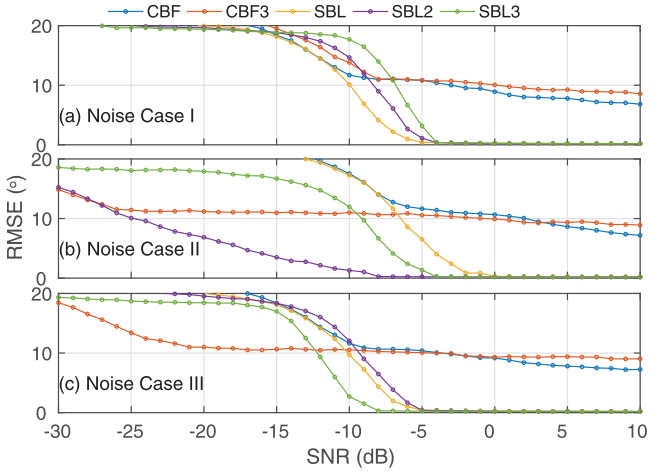

Fig. 13. RMSE vs. SNR with three off-grid sources at $\{-2.889^\circ, 2.222^\circ, 50.178^\circ\}$ and same power {10, 10, 10} dB.

5.4.4. SBL Comparison

Here, we compare to other SBL technique [39, Eq. (18)], see Fig. 11. Noise estimates and all parameters are identical to those used in Fig. 6. In comparison, the two SBL algorithms give similar results, demonstrating that the noise estimate plays a more significant role than the SBL technique.

5.4.5. Off-grid and random sources

We are here concerned with a scenario where the ideal conditions are not satisfied by simulating off-grid sources with either different powers for each DOA (Fig. 12) or identical powers (Fig. 13). For identical powers the solution here diverges at a lower array SNR, though the divergence is then faster.

We also investigate when all 3 DOAs are uniformly random, subject to a DOA separation $|\sin\theta_i - \sin\theta_j| > 0.05$ corresponding to at least 2.5° (computed at broadside, at endfire it is 18°), see Fig. 14. Naturally each SBL that satisfies the noise assumptions does well (SBL for Noise Case I, SBL2 for Noise Case II, and SBL3 for Noise Case III).

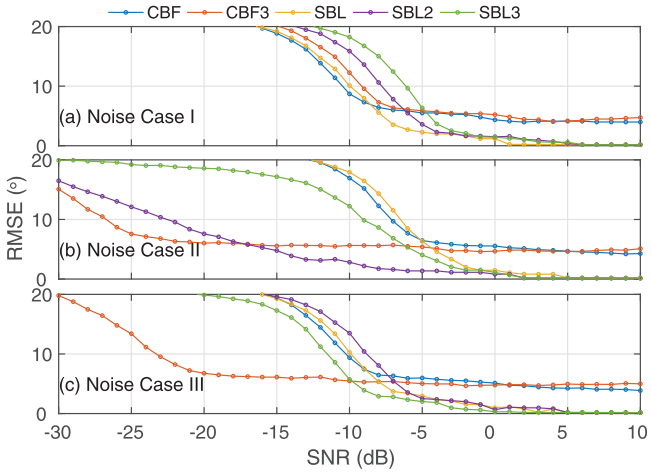

Fig. 14. RMSE vs. SNR with three sources at uniform random DOAs separated by at least 2.5° and the same power {10, 10, 10} dB.

The 2.5° separation is so small that CBF and CBF3 cannot resolve the DOAs even at high SNR. At very low SNR CBF3, which is whitened to only retain phase, performs better than any other of the processors. For larger separation (not shown) where CBF3 does not have resolution problems it performs even better that any of the other processors at low SNR. This supports the empirical evidence that retaining phase when extracting information from noise is beneficial [23],[34].

6. Conclusion

Stochastic likelihood based DOA estimators from long-term observations corrupted by non-stationary additive noise are proposed and discussed. In such a setting, the DOA estimators which are developed for stationary noise perform poorly, meaning the DOA estimation fails at too high SNR. Therefore a heteroscedastic Gaussian noise model is introduced with the noise variance varying across sensors and snapshots. We develop sparse Bayesian learning (SBL) approaches to estimate source powers, source DOAs, and the noise variance parameters.

We considered three noise models as follows. Noise Case I: noise variance is assumed constant, the classic homoscedastic model. Noise Case II: noise variance is varying across snapshots. Noise Case III: noise variance is varying across snapshots and sensors, fully heteroscedastic model. These can be seen as nested models, with Noise Case I being the reduced model and Noise Case III the full model. Multiple hypotheses testing could provide a formal tool for model selection based on the data.

A more practical approach is to observe which model gives highest goodness of fit or lowest RMSE. For the simulations performed there are significant gains in choosing a noise model with more variance parameters. The simulations indicate that the proposed SBL algorithm tailored to Noise Case III ("SBL3") performs satisfactorily for all three noise cases. The loss in RMSE is negligible when SBL3 is used for array data following Noise Cases I or II. We observe that SBL3 pushes the SNR threshold (where DOA estimation fails) down to −30 dB when the array data follow Noise Case III and only a single source is present. When multiple sources are present, this qualitative behavior of the proposed SBL algorithms remains unchanged at low SNR.

Simulations show that normalizing the array data magnitude (such that only the phase information is retained) is simple and useful for estimating well separated DOAs in heteroscedastic Gaussian noise. For the estimation of (multiple) closely spaced sources, a problem specific SBL approach gives much lower RMSE in the DOA estimates.

The processing has been demonstrated on experimental data [62].

Acknowledgment

Supported by the Office of Naval Research, Grant No. N00014-18-1-2118.

References

[1] K.P. Murphy, Machine learning: a probabilistic perspective, MIT press, 2012.
[2] C.M. Bishop, Pattern recognition and machine learning, Springer, 2006.
[3] A.M. Zoubir, V. Koivunen, E. Ollila, M. Muma, Robust statistics for signal processing, Cambridge University Press, 2018.
[4] P.J. Huber, Robust statistics, Springer, 2011.
[5] R. Maronna, D. Martin, V. Yohai, Robust statistics, John Wiley, Chichester., 2006.
[6] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, Econometrica (1982) 987–1007.
[7] T. Thai, R. Cogranne, F. Retraint, Camera model identification based on the heteroscedastic noise model, IEEE trans, Image Process 23 (1) (2014) 250–263.
[8] M. Viberg, P. Stoica, B. Ottersten, Maximum likelihood array processing in spatially correlated noise fields using parameterized signals, IEEE Trans Signal Process 45 (4) (1997) 996–1004.
[9] M. Pesavento, A.B. Gershman, Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise, IEEE Trans Signal Process 49 (7) (2001) 1310–1324.
[10] C.E. Chen, F. Lorenzelli, R. Hudson, K. Yao, Stochastic maximum-likelihood doa estimation in the presence of unknown nonuniform noise, IEEE Trans. Signal Process 56 (7) (2008) 3038–3044.
[11] T. Li, A. Nehorai, Maximum likelihood direction finding in spatially colored noise fields using sparse sensor arrays, IEEE Trans Signal Process 59 (3) (2011) 1048–1062.
[12] A. Gholipour, B. Zakeri, K. Mafinezhad, Non-stationary additive noise modelling in direction-of-arrival estimation, IET Commun. 10 (15) (2016) 2054–2059.
[13] M. Hurtado, C. Muravchik, A. Nehorai, Enhanced sparse Bayesian learning via statistical thresholding for signals in structured noise, IEEE Trans. Signal Process 61 (21) (2013) 5430–5443.
[14] S. Pazos, M. Hurtado, C. Muravchik, On sparse methods for array signal processing in the presence of interference, IEEE Ant. Wireless Prop. Lett. 14 (2015) 1165–1168.
[15] J. Dai, H.C. So, Sparse Bayesian learning approach for outlier-resistant direction-of-arrival estimation, IEEE Trans. Signal Process. 66 (3) (2018) 744–756.
[16] H. Amiri, H. Amindavar, R. Kirlin, Array signal processing using garch noise modeling, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04), volume 2, IEEE, 2004, pp. ii–105.
[17] H. Amiri, H. Amindavar, M. Kamarei, Underwater noise modeling and direction-finding based on heteroscedastic time series, EURASIP J Adv Signal Process 2007 (1) (2007) 071528, doi:10.1155/2007/71528.
[18] H. Cox, Line array performance when the signal coherence is spatially dependent, J Acoust. Soc. Am. 54 (6) (1973) 1743–1746.
[19] A. Paulraj, T. Kailath, Direction of arrival estimation by eigenstructure methods with imperfect spatial coherence of wave fronts, J Acoust. Soc. Am. 83 (3) (1988) 1034–1040.
[20] R. Lefort, R. Emmetière, S. Bourmani, G. Real, A. Drémeau, Sub-antenna processing for coherence loss in underwater direction of arrival estimation, J Acoust. Soc. Am. 142 (4) (2017) 2143–2154.
[21] N. Law, C.D. Mackay, J. Baldwin, Lucky imaging: high angular resolution imaging in the visible from the ground, Astronomy & Astrophysics 446 (2) (2006) 739–745.
[22] H. Ge, I.P. Kirsteins, Lucky ranging with towed arrays in underwater environments subject to non-stationary spatial coherence loss, in: IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP), 2016, pp. 3156–3160, doi:10.1109/ICASSP.2016.7472259.
[23] P. Gerstoft, M.C. Fehler, K.G. Sabra, When katrina hit california, Geophys. Res. Lett. 33 (17) (2006) L17308.
[24] P. Gerstoft, P.D. Bromirski, "Weather bomb" induced seismic signals, Science 353 (6302) (2016) 869–870.
[25] P. Roux, K.G. Sabra, P. Gerstoft, W.A. Kuperman, M.C. Fehler, P-Waves from cross-correlation of seismic noise, Geophys. Res. Lett. 32 (19) (2005) L19303.
[26] N. Harmon, P. Gerstoft, C.A. Rychert, G.A. Abers, M.S. de La Cruz, K.M. Fischer, Phase velocities from seismic noise using beamforming and cross correlation in costa rica and nicaragua, Geophy. Res. Lett. 35 (19) (2008) 19303.
[27] M. Landès, F. Hubans, N.M. Shapiro, A. Paul, M. Campillo, Origin of deep ocean microseisms by using teleseismic body waves, J. Geophys Res: Solid Earth 115 (B5) (2010) B05302.
[28] Z. Zhan, S. Ni, D.V. Helmberger, R.W. Clayton, Retrieval of moho-reflected shear wave arrivals from ambient seismic noise, Geophys. J. Int. 182 (1) (2010) 408–420.
[29] C. Weemstra, W. Westra, R. Snieder, L. Boschi, On estimating attenuation from the amplitude of the spectrally whitened ambient seismic field, Geophys. J. Int. 197 (3) (2014) 1770–1788, doi:10.1093/gji/ggu088.

[30] P. Gerstoft, W.S. Hodgkiss, M. Siderius, C.F. Huang, C. Harrison, Passive fathometer processing, J. Acoust. Soc. Am. 123 (3) (2008) 1297–1305.
[31] K. Sabra, D. Dowling, Blind deconvolution in ocean waveguides using artificial time reversal, J. Acoust. Soc. Am. 116 (1) (2004) 262–271.
[32] K. Sabra, H.C. Song, D. Dowling, Ray-based blind deconvolution in ocean sound channels, J. Acoust. Soc. Am. 127 (2) (2010) EL42–EL47.
[33] O. Schwartz, S. Gannot, Speaker tracking using recursive em algorithms, IEEE Trans. Audio, Speech, Language Process. 22 (2) (2014) 392–402.
[34] Y. Dorfan, S. Gannot, Tree-based recursive expectation-maximization algorithm for localization of acoustic sources, IEEE Trans. Audio, Speech and Lang Process. 23 (10) (2015) 1692–1703.
[35] N. Riahi, P. Gerstoft, The seismic traffic footprint: tracking trains, aircraft, and cars seismically, Geophys. Res. Lett. 42 (8) (2015) 2674–2681.
[36] H.V. Trees, Optimum array processing, Wiley-Interscience, New York, 2002. Ch. 1–10.
[37] 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.
[38] A. Xenaki, P. Gerstoft, K. Mosegaard, Compressive beamforming, J. Acoust. Soc. Am. 136 (1) (2014) 260–271.
[39] D.P. Wipf, B. Rao, An empirical Bayesian strategy for solving the simultaneous sparse approximation problem, IEEE Trans. Signal Process. 55 (7) (2007) 3704–3716.
[40] D.P. Wipf, S. Nagarajan, Beamforming using the relevance vector machine, in: Proc. 24th Int. Conf. Machine Learning, 2007, pp. 1–4, doi:10.1145/1273496.1273625. New York, NY, USA.
[41] P. Gerstoft, A. Xenaki, C. Mecklenbräuker, Multiple and single snapshot compressive beamforming, J. Acoust. Soc. Am. 138 (4) (2015) 2003–2014.
[42] P. Gerstoft, C.F. Mecklenbräuker, A. Xenaki, S. Nannuru, Multisnapshot sparse Bayesian learning for DOA, IEEE Signal Process. Lett. 23 (10) (2016) 1469–1473.
[43] S. Nannuru, K.L. Gemba, P. Gerstoft, W.S. Hodgkiss, C.F. Mecklenbräuker, Sparse Bayesian learning with multiple dictionaries, Signal Processing 159 (2019) 159–170, doi:10.1016/j.sigpro.2019.02.003.
[44] P. Stoica, P. Babu, SPICE And LIKES: two hyperparameter-free methods for sparse-parameter estimation, Signal Process. 92 (7) (2012) 1580–1590.
[45] D.P. Wipf, B. Rao, Sparse Bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164.
[46] Z. Zhang, B.D. Rao, Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning, IEEE J Sel. Topics Signal Process. 5 (5) (2011) 912–926.
[47] 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, doi:10.1109/TWC.2012.090312.111912.
[48] J. Zhang, Z. Chen, P. Cheng, X. Huang, Multiple-measurement vector based implementation for single-measurement vector sparse Bayesian learning with reduced complexity, Signal Process. 118 (2016) 153–158.
[49] 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, doi:10.1109/TSP.2016.2546231.
[50] J. Böhme, Source-parameter estimation by approximate maximum likelihood and nonlinear regression, IEEE J. Oceanic Eng. 10 (3) (1985) 206–212, doi:10.1109/JOE.1985.1145098.
[51] A. Jaffer, Maximum likelihood direction finding of stochastic sources: A separable solution, in: IEEE Int. Conf. on Acoust., Speech, and Sig. Process. (ICASSP-88), volume 5, 1988, pp. 2893–2896, doi:10.1109/ICASSP.1988.197258.
[52] P. Stoica, A. Nehorai, On the concentrated stochastic likelihood function in array processing, Circuits Syst. Signal Process. 14 (5) (1995) 669–674.
[53] F. Athley, Threshold region performance of maximum likelihood direction of arrival estimators, IEEE Trans Signal Process. 53 (4) (2005) 1359–1373.
[54] J.F. Böhme, Estimation of spectral parameters of correlated signals in wavefields, Signal Process 11 (4) (1986) 329–337.
[55] H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach, IEEE Signal Process. Mag. 13 (4) (1996) 67–94.
[56] P. Stoica, B. Ottersten, M. Viberg, R. Moses, Maximum likelihood array processing for stochastic coherent sources, IEEE Trans. Signal Process. 44 (1) (1996) 96–105.
[57] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Machine Learning Research 1 (2001) 211–244.
[58] D. Kraus, Approximative maximum-likelihood-schätzung und verwandte verfahren zur ortung und signalschätzung mit sensorgruppen, Shaker Verlag, Aachen, Germany, 1993.
[59] G. Bienvenu, L. Kopp, Optimality of high resolution array processing using the eigensystem approach, IEEE Trans on Acous., Speech, Signal Process. 31 (5) (1983) 1235–1248, doi:10.1109/TASSP.1983.1164185.
[60] E. Ollila, Multichannel sparse recovery of complex-valued signals using Huber's criterion, in: 3rd Int. Workshop on Compressed Sensing Theory and Appl. to Radar, Sonar, and Remote Sensing, Pisa, Italy, 2015.
[61] Z. Zhang, T.P. Jung, S. Makeig, Z. Pi, B. Rao, Spatiotemporal sparse Bayesian learning with applications to compressed sensing of multichannel physiological signals, IEEE Trans. Neural Syst. and Rehab. Eng. 22 (6) (2014) 1186–1197, doi:10.1109/TNSRE.2014.2319334.
[62] K.L. Gemba, S. Nannuru, P. Gerstoft, Robust ocean acoustic localization with sparse Bayesian learning, IEEE Journal of Selected Topics in Signal Processing. doi:10.1109/JSTSP.2019.2900912.