

DOA ESTIMATION IN HETEROSCEDASTIC NOISE WITH SPARSE BAYESIAN LEARNING

Peter Gerstoft* Santosh Nannuru* Christoph F. Mecklenbräuer† and Geert Leus⁺

*University of California San Diego

†Inst. of Telecommunications, TU Wien, 1040 Wien, Austria

⁺Dept. of Electrical Eng., Math. and Comp. Science, Delft Univ. of Technology

ABSTRACT

The paper considers direction of arrival (DOA) estimation from long-term observations in a noisy environment. In such an environment the noise source might evolve, causing the stationary models to fail. Therefore a heteroscedastic Gaussian noise model is introduced where the variance can vary across observations and sensors. The source amplitudes are assumed independent zero-mean complex Gaussian distributed with unknown variances (i.e. the source powers), leading to 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. Simulations demonstrate that taking the heteroscedastic noise into account improves DOA estimation.

Index Terms— DOA estimation, sparsity, sparse Bayesian learning, heteroscedastic noise

1. INTRODUCTION

With long observation times weak signals can be extracted in a noisy environment. Most analytic treatments analyze these cases assuming Gaussian noise with constant variance. For long observation times the noise process though is likely to change with time causing the noise variance to evolve. This is called a heteroscedastic Gaussian process, meaning that the noise variance is evolving. While the noise variance is a nuisance parameter that we are not interested in, it still needs to be estimated or included in the processing in order to obtain an accurate estimate of the weaker signals.

Accounting for the noise variation is certainly important for machine learning [1] and related to robust statistics [2]. In statistical signal processing, the noise has been assumed to vary spatially [3, 4, 5], but spatiotemporally varying noise considered here has not been studied. The proposed processing could be applied to spatial coherence loss [6, 7, 8] or to wavefront decorrelation, where turbulence causes the wave front to be incoherent for certain observations (thus more

noisy). This has led to so-called lucky imaging in astronomy or lucky ranging in ocean acoustics [9], where only the measurements giving good results are used. 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 [10, 11, 12, 13] using multiple measurement vectors (MMV or multiple snapshot). We solve the MMV problem using the sparse Bayesian learning (SBL) framework [12, 14, 15].

We base our development on our fast SBL method [14, 15] 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 the number observations. Existing techniques are based on minimization-majorization [16] and expectation maximization (EM) [12, 17, 18, 19, 20, 21, 22], though not all estimates work well. Instead, we estimate the unknown variances using approximate stochastic ML [23, 24, 25] modified to obtain noise estimates even for a single observation. Further details can be found in the full paper [26]

1.1. Heteroscedastic observation model

For the l th observation snapshot, we assume the linear model

$$\mathbf{y}_l = \mathbf{A}\mathbf{x}_l + \mathbf{n}_l, \quad (1)$$

where the dictionary $\mathbf{A} \in \mathbb{C}^{N \times M}$ is constant and known and the source vector $\mathbf{x}_l \in \mathbb{C}^M$ contains the physical information of interest. Further, $\mathbf{n}_l \in \mathbb{C}^N$ is additive zero-mean circularly symmetric complex Gaussian noise, which is generated from a *heteroscedastic* Gaussian process $\mathbf{n}_l \sim \mathcal{CN}(\mathbf{n}_l; \mathbf{0}, \mathbf{\Sigma}_{\mathbf{n}_l})$. We assume that the covariance matrix is diagonal and parameterized as

$$\mathbf{\Sigma}_{\mathbf{n}_l} = \sum_{n=1}^N \sigma_{n,l}^2 \mathbf{J}_n = \text{diag}(\sigma_{1,l}^2, \dots, \sigma_{N,l}^2), \quad (2)$$

where $\mathbf{J}_n = \text{diag}(\mathbf{e}_n) = \mathbf{e}_n \mathbf{e}_n^T$ with \mathbf{e}_n the n th standard basis vector. Note that the covariance matrices $\mathbf{\Sigma}_{\mathbf{n}_l}$ are varying

Supported by the Office of Naval Research, Grant Nos. N00014-1110439 and FTW Austria's "Compressed channel state information feedback for time-variant MIMO channels".

over the snapshot index $l = 1, \dots, L$. We consider three cases for the a priori knowledge on the noise covariance model (2):

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

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, $\sigma_{n,l}^2 = \sigma_{0,l}^2$ and it varies over snapshots. The noise variance is heteroscedastic in time (across snapshots).

III: No additional constraints other than (2). The noise variance is heteroscedastic across both time and space (sensors and snapshots.)

1.2. Array model

Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_L] \in \mathbb{C}^{M \times L}$ be the complex source amplitudes, $x_{ml} = [\mathbf{X}]_{m,l} = [\mathbf{x}_l]_m$ with $m \in \{1, \dots, M\}$ and $l \in \{1, \dots, L\}$, at M DOAs (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 $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L] \in \mathbb{C}^{N \times L}$. A linear regression model relates the array data \mathbf{Y} to the source amplitudes \mathbf{X} as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N}. \quad (3)$$

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

We assume $M > N$ and thus (3) is underdetermined. In the presence of only few stationary sources, the source vector \mathbf{x}_l is K -sparse with $K \ll M$. We define the l th active set

$$\mathcal{M}_l = \{m \in \mathbb{N} | x_{ml} \neq 0\}, \quad (4)$$

and assume $\mathcal{M}_l = \mathcal{M} = \{m_1, \dots, m_K\}$ 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} .

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 , $m = 1, \dots, M$,

$$p(x_{ml}; \gamma_m) = \begin{cases} \delta(x_{ml}), & \text{for } \gamma_m = 0 \\ \frac{1}{\pi\gamma_m} e^{-|x_{ml}|^2/\gamma_m}, & \text{for } \gamma_m > 0 \end{cases}, \quad (5)$$

$$p(\mathbf{X}; \boldsymbol{\gamma}) = \prod_{l=1}^L \prod_{m=1}^M p(x_{ml}; \gamma_m) = \prod_{l=1}^L \mathcal{CN}(\mathbf{x}_l; \mathbf{0}, \boldsymbol{\Gamma}), \quad (6)$$

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

$$\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma}) = \mathbb{E}[\mathbf{x}_l \mathbf{x}_l^H; \boldsymbol{\gamma}], \quad (7)$$

as $\text{rank}(\boldsymbol{\Gamma}) = \text{card}(\mathcal{M}) = K \leq M$ (typically $K \ll M$). Note that the diagonal elements of $\boldsymbol{\Gamma}$, i.e., $\boldsymbol{\gamma} \geq \mathbf{0}$, represent source

powers. When the variance $\gamma_m = 0$, then $x_{ml} = 0$ with probability 1.

1.3. Stochastic likelihood

We here derive the well-known stochastic likelihood function [27, 28, 29]. Given the linear model (3) with Gaussian source (6) and noise (2) the array data \mathbf{Y} is Gaussian with for each snapshot l the covariance $\boldsymbol{\Sigma}_{\mathbf{y}_l}$ given by

$$\boldsymbol{\Sigma}_{\mathbf{y}_l} = \mathbb{E}[\mathbf{y}_l \mathbf{y}_l^H] = \boldsymbol{\Sigma}_{\mathbf{n}_l} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^H \quad (8)$$

$$p(\mathbf{Y}) = \prod_{l=1}^L \mathcal{CN}(\mathbf{y}_l; \mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}_l}) = \prod_{l=1}^L \frac{e^{-\mathbf{y}_l^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{y}_l}}{\pi^N \det \boldsymbol{\Sigma}_{\mathbf{y}_l}}, \quad (9)$$

The L -snapshot log-likelihood for estimating $\boldsymbol{\gamma}$ and $\boldsymbol{\Sigma}_{\mathbf{N}} = \{\boldsymbol{\Sigma}_{\mathbf{n}_1}, \dots, \boldsymbol{\Sigma}_{\mathbf{n}_L}\}$ is

$$\log p(\mathbf{Y} | \boldsymbol{\gamma}, \boldsymbol{\Sigma}_{\mathbf{N}}) \propto - \sum_{l=1}^L (\mathbf{y}_l^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{y}_l + \log \det \boldsymbol{\Sigma}_{\mathbf{y}_l}). \quad (10)$$

This likelihood function is identical to the Type II likelihood function (evidence) in standard SBL [18, 17, 14] which is obtained by treating $\boldsymbol{\gamma}$ as a hyperparameter. The parameter estimates $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{N}}$ are obtained by maximizing the likelihood,

$$(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\Sigma}}_{\mathbf{N}}) = \arg \max_{\boldsymbol{\gamma} \geq \mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{N}}} \log p(\mathbf{Y}; \boldsymbol{\gamma}, \boldsymbol{\Sigma}_{\mathbf{N}}). \quad (11)$$

The likelihood function (10) is similar to the ones derived for SBL and LIKES [16]. The goal is thus to solve (11) and the active DOAs \mathcal{M} is where $\hat{\boldsymbol{\gamma}} > 0$.

1.4. Source power estimation

Let us now focus on the SBL algorithm solving (11). The algorithm iterates between the source power estimates $\hat{\boldsymbol{\gamma}}$ derived in this section and the noise variance estimates $\hat{\boldsymbol{\Sigma}}_{\mathbf{N}}$ computed in Sec. 2. We impose the diagonal structure $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma})$, in agreement with (6), and form derivatives of (10) with respect to the diagonal elements γ_m , cf. [27]. The derivative of (10) is

$$\begin{aligned} \frac{\partial \log p(\mathbf{Y}; \boldsymbol{\gamma}, \boldsymbol{\Sigma}_{\mathbf{N}})}{\partial \gamma_m} &= \sum_{l=1}^L (\mathbf{a}_m^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{y}_l \mathbf{y}_l^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{a}_m - \mathbf{a}_m^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{a}_m) \\ &= \sum_{l=1}^L \mathbf{a}_m^H (\boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{y}_l \mathbf{y}_l^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} - \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1}) \mathbf{a}_m \quad (12) \\ &= \sum_{l=1}^L |\mathbf{y}_l^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{a}_m|^2 - \sum_{l=1}^L \mathbf{a}_m^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{a}_m. \quad (13) \end{aligned}$$

We impose the necessary condition $\frac{\partial \log p(\mathbf{Y}; \boldsymbol{\gamma}, \boldsymbol{\Sigma}_{\mathbf{N}})}{\partial \gamma_m} = 0$ for solving (11). Assuming $\boldsymbol{\gamma}_m^{\text{old}}$ and $\boldsymbol{\Sigma}_{\mathbf{y}_l}$ given (from previous

iterations) and forcing (13) to zero, we obtain the following fixed point iteration [30] for the γ_m [14]:

$$\gamma_m^{\text{new}} = \gamma_m^{\text{old}} \left(\frac{\sum_{l=1}^L |\mathbf{y}_l^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{a}_m|^2}{\sum_{l=1}^L \mathbf{a}_m^H \boldsymbol{\Sigma}_{\mathbf{y}_l}^{-1} \mathbf{a}_m} \right)^b. \quad (14)$$

It is not clear what value of b to use. Higher b -values give faster convergence to an estimate, but we have not tested for optimal values and this value also depends on other factors such as which noise estimate is used. A value of $b = 1$ gives the update equation used in [31, 12] and $b = 0.5$ gives the update equation used in [14].

2. NOISE VARIANCE ESTIMATION

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. In SBL, the noise variance controls the sharpness of the peaks in the $\boldsymbol{\gamma}$ spectrum, with higher noise levels giving broader peaks.

We estimate the noise variance for the three noise cases in Sec. 1.1. In Secs. 2.1–2.3, we will assume the support of $\boldsymbol{\gamma}$ is known. In simulations we estimate support from peaks of $\boldsymbol{\gamma}$.

2.1. Noise estimate, Case I

Under Noise Case I, where $\boldsymbol{\Sigma}_{\mathbf{n}_l} = \sigma^2 \mathbf{I}_N$ with \mathbf{I}_N the identity matrix of size N , stochastic ML [20, 23, 25] provides an asymptotically efficient estimate of σ^2 if the set of active DOAs \mathcal{M} is known.

Let $\boldsymbol{\Gamma}_{\mathcal{M}} = \text{diag}(\boldsymbol{\gamma}_{\mathcal{M}}^{\text{new}})$ be the covariance matrix of the K active sources obtained above with corresponding active steering matrix $\mathbf{A}_{\mathcal{M}}$ which maximizes (10). The corresponding data covariance matrix is

$$\boldsymbol{\Sigma}_{\mathbf{y}_l} = \sigma^2 \mathbf{I}_N + \mathbf{A}_{\mathcal{M}} \boldsymbol{\Gamma}_{\mathcal{M}} \mathbf{A}_{\mathcal{M}}^H. \quad (15)$$

Note that for Noise Case I, the data covariance matrices (8) and (15) are identical. Let us then define the projection matrix onto the subspace spanned by the active steering vectors

$$\mathbf{P} = \mathbf{A}_{\mathcal{M}} \mathbf{A}_{\mathcal{M}}^+ = \mathbf{A}_{\mathcal{M}} (\mathbf{A}_{\mathcal{M}}^H \mathbf{A}_{\mathcal{M}})^{-1} \mathbf{A}_{\mathcal{M}}^H = \mathbf{P}^H = \mathbf{P}^2 \quad (16)$$

and the sample covariance matrix

$$\mathbf{S}_{\mathbf{y}} = \frac{1}{L} \mathbf{Y} \mathbf{Y}^H. \quad (17)$$

We use the approximate ML noise variance estimate $\hat{\sigma}^2$ obtained by [23, 25]

$$\sigma^2 = \frac{\text{tr}[\boldsymbol{\Sigma}_{\mathbf{y}_l} - \mathbf{P} \mathbf{S}_{\mathbf{y}}]}{N - K} = \frac{\text{tr}[(\mathbf{S}_{\mathbf{y}} - \mathbf{P} \mathbf{S}_{\mathbf{y}}) + \boldsymbol{\epsilon}]}{N - K} \quad (18)$$

$$\approx \frac{\text{tr}[(\mathbf{I}_N - \mathbf{P}) \mathbf{S}_{\mathbf{y}}]}{N - K} = \hat{\sigma}^2. \quad (19)$$

Here, we defined the power estimate error $\boldsymbol{\epsilon} = \text{tr}[\boldsymbol{\Sigma}_{\mathbf{y}_l} - \mathbf{S}_{\mathbf{y}}]$.

The above approximation motivates the noise power estimate for Noise Case I (19), which is *error-free* if $\text{tr}[\boldsymbol{\Sigma}_{\mathbf{y}}] = \text{tr}[\mathbf{S}_{\mathbf{y}}]$, *unbiased* because $\text{E}[\boldsymbol{\epsilon}] = 0$, *consistent* since also its variance tends to zero for $L \rightarrow \infty$, and *asymptotically efficient* as it approaches the CRLB for $L \rightarrow \infty$. The Noise Case I estimate (19) is valid even for one snapshot.

2.2. Noise estimate, Case II

For Noise Case II, where $\boldsymbol{\Sigma}_l = \sigma_l^2 \mathbf{I}_N$, we apply (19) for each snapshot l individually, leading to

$$\hat{\sigma}_l^2 = \frac{\text{tr}[(\mathbf{I}_N - \mathbf{P}) \mathbf{y}_l \mathbf{y}_l^H]}{N - K} = \frac{\|(\mathbf{I}_N - \mathbf{P}) \mathbf{y}_l\|_2^2}{N - K}. \quad (20)$$

2.3. Noise estimate, Case III

Let us start from the definition of the noise covariance

$$\boldsymbol{\Sigma}_{\mathbf{n}_l} = \text{diag}[\sigma_{1,l}^2, \dots, \sigma_{n,l}^2, \dots, \sigma_{N,l}^2] \quad (21)$$

$$\begin{aligned} &= \text{E}[(\mathbf{y}_l - \mathbf{A} \mathbf{x}_l)(\mathbf{y}_l - \mathbf{A} \mathbf{x}_l)^H] \\ &= \text{E}[(\mathbf{y}_l - \mathbf{A}_{\mathcal{M}} \mathbf{x}_{\mathcal{M},l})(\mathbf{y}_l - \mathbf{A}_{\mathcal{M}} \mathbf{x}_{\mathcal{M},l})^H] \end{aligned} \quad (22)$$

This motivates plugging-in the single-observation signal estimate $\hat{\mathbf{x}}_{\mathcal{M},l} = \mathbf{A}_{\mathcal{M}}^+ \mathbf{y}_l \in \mathbb{C}^K$ for the active (non-zero) entries in \mathbf{x}_l . This estimate is based on the single observation \mathbf{y}_l and the projection matrix (16), giving the rank-1 estimate

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{n}_l} = (\mathbf{I} - \mathbf{P}) \mathbf{y}_l \mathbf{y}_l^H (\mathbf{I} - \mathbf{P}). \quad (23)$$

Since the signal estimate $\hat{\mathbf{x}}_l$ maximizes the estimated signal power, this noise covariance estimate is biased and the noise level is likely underestimated.

Since we assume the noise independent across sensors, all off-diagonal elements of $\boldsymbol{\Sigma}_{\mathbf{n}_l}$ are known to be zero. With this constraint in mind, we modify (23) as

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{n}_l} = \text{diag}[(\mathbf{I} - \mathbf{P}) \mathbf{y}_l \mathbf{y}_l^H (\mathbf{I} - \mathbf{P})]. \quad (24)$$

The estimate (24) 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 $\hat{\boldsymbol{\Sigma}}_{\mathbf{n}_l}$ in (23) is not invertible whereas the diagonal constraint in (24) leads to a non-singular estimate of $\boldsymbol{\Sigma}_{\mathbf{n}_l}$ with high probability (it is singular only if an element of \mathbf{y}_l is 0). As a result, the expression for $\boldsymbol{\Sigma}_{\mathbf{y}_l}$ that is used for estimating $\boldsymbol{\gamma}$ in (14) is likely invertible.

On the other hand, an overestimate of the noise is easily obtained by assuming $\mathbf{x}_l = \mathbf{0}$ which is equivalent to setting $\mathbf{P} = \mathbf{0}$ in (24), resulting in

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{n}_l} = \text{diag}[\mathbf{y}_l \mathbf{y}_l^H], \quad \hat{\sigma}_{n,l} = |y_{nl}|. \quad (25)$$

This can be shown by the stochastic ML estimate for no sources ($\mathcal{M} = 0$) or very low power sources.

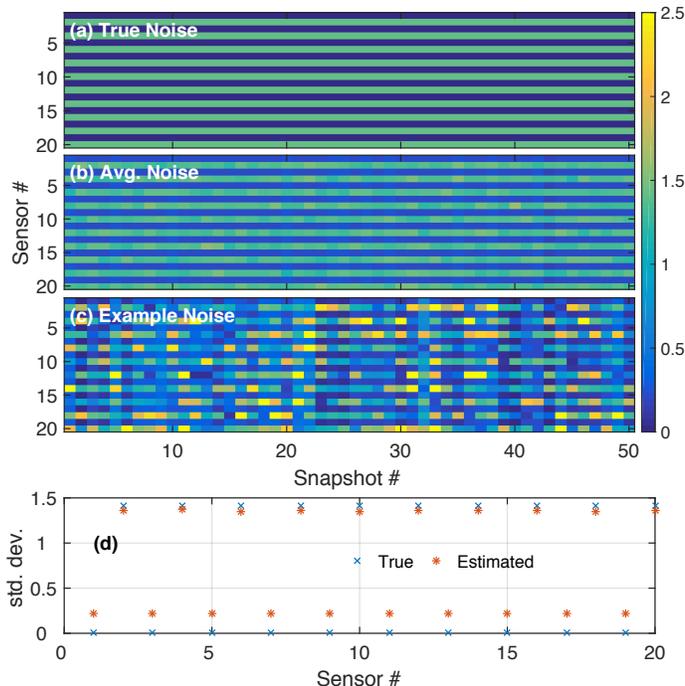


Fig. 1. Single source at -3° , array SNR = 0 dB, SBL3 noise standard deviation estimate statistics: (a) true, (b) average estimated (100 simulations), (c) a typical estimate, and (d) average across simulations and snapshots.

3. EXAMPLES

An example statistic of the heteroscedastic noise standard deviation is shown in Fig. 1 for a 20 element array with a single source. The standard deviation for each sensor is either 0 or $\sqrt{2}$ (Fig. 1a). The estimates of the standard deviation are obtained from (24) (Fig. 1b, 1c). Average of estimated noise (Fig. 1b) resembles well the true noise (Fig. 1a) whereas the sample standard deviation estimate (Fig. 1c) has high variability—each estimate is based on just observation. Given many simulations and snapshots, however, the mean of the estimated standard deviation is close to the true noise (Fig. 1d).

In the analysis of seismic data the noise for each snapshot was observed to be log-normal distributed [32]. In the simulations, the noise follows a normal-uniform hierarchical model. The noise is complex zero-mean Gaussian with the standard deviation uniformly distributed over two decades, i.e., $\log_{10}\sigma_{n,l} \sim \mathcal{U}(-1,1)$, where \mathcal{U} is the uniform distribution. Three noise cases are simulated:

- (a) Noise Case I: constant noise standard deviation over snapshots and sensors,
- (b) Noise Case II: standard deviation changes across snapshots with $\log_{10}\sigma_l \sim \mathcal{U}(-1,1)$, and
- (c) Noise Case III: standard deviation changes across both snapshots and sensors with $\log_{10}\sigma_{n,l} \sim \mathcal{U}(-1,1)$.

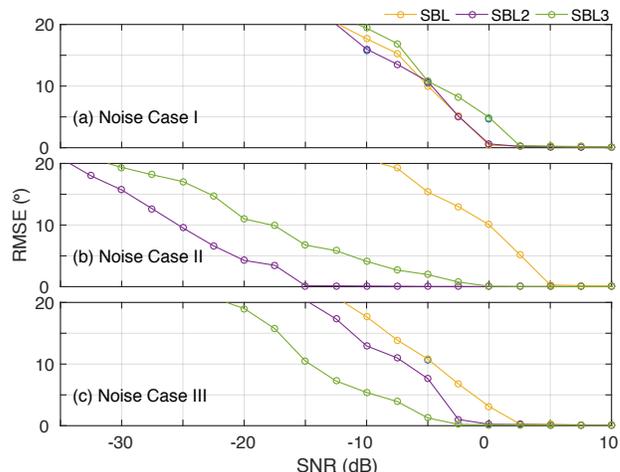


Fig. 2. RMSE vs. SNR with the three sources at $\{-3^\circ, 2^\circ, 50^\circ\}$ and power $\{10, 22, 20\}$ dB.

In example 2 (Fig. 2), e consider three sources located at $[-3, 2, 50]^\circ$ with power $[10, 22, 20]$ dB. The complex source amplitude is stochastic and there is additive heteroscedastic Gaussian noise with SNR variation from -35 to 10 dB. The $N=20$ elements sensor array with half-wavelength spacing observe $L=50$ 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 $\text{SNR} = 10 \log_{10} [\mathbb{E}\{\|\mathbf{A}\mathbf{x}_l\|_2^2\} / \mathbb{E}\{\|\mathbf{n}_l\|_2^2\}]$. The root mean squared error (RMSE) of the DOA estimates over 100 noise realizations is used for evaluating the algorithms. We use three SBL methods with $\boldsymbol{\gamma}$ update (14), namely; *SBL*: Noise Case I, standard SBL, with σ from (19); *SBL2*: Noise Case II, with σ_l from (20); *SBL3*: Noise Case III, with $\sigma_{n,l}$ from (24).

For Noise Case I (Fig. 2a), standard SBL is optimal, but SBL2 and SBL3 perform similar. For Noise Case II (Fig. 2b), SBL2 is optimal and fails 15 dB later than SBL3 and 20 dB later than SBL. For Noise Case III (Fig. 2c), SBL3 is optimal and fails 3 dB later than SBL2 and 8 dB later than SBL. Thus, the simulation demonstrates that estimating the noise carefully gives improved DOA estimation at low SNR.

4. CONCLUSION

Stochastic likelihood based methods for DOA estimation from observations corrupted by nonstationary additive noise is discussed. In such setting estimators based on a stationary noise model perform poorly. A heteroscedastic Gaussian noise model is introduced with noise variance varying across sensors and snapshots. We develop SBL approaches to estimate heteroscedastic Gaussian noise parameters. Using a problem specific SBL approach gives much lower DOA RMSE.

5. REFERENCES

- [1] K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [2] P. J. Huber. *Robust statistics*. Springer, 2011.
- [3] M. Viberg, P. Stoica, and B. Ottersten. Maximum likelihood array processing in spatially correlated noise fields using parameterized signals. *IEEE Trans Signal Proc.*, 45(4):996–1004, 1997.
- [4] C. E. Chen, F. Lorenzelli, R.E Hudson, and K. Yao. Stochastic maximum-likelihood doa estimation in the presence of unknown nonuniform noise. *IEEE Trans. Signal Proc.*, 56(7):3038–3044, 2008.
- [5] T Li and A. Nehorai. Maximum likelihood direction finding in spatially colored noise fields using sparse sensor arrays. *IEEE Trans Signal Proc.*, 59(3):1048–1062, 2011.
- [6] H. Cox. Line array performance when the signal coherence is spatially dependent. *J Acoust. Soc. Am.*, 54(6):1743–1746, 1973.
- [7] A. Paulraj and T. Kailath. Direction of arrival estimation by eigenstructure methods with imperfect spatial coherence of wave fronts. *J Acoust. Soc. Am.*, 83(3):1034–1040, 1988.
- [8] R. Lefort, R. Emmeti re, S. Bourmani, G. Real, and A. Dr meau. Sub-antenna processing for coherence loss in underwater direction of arrival estimation. *J Acoust. Soc. Am.*, 142(4):2143–2154, 2017.
- [9] H. Ge and 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 Proc. (ICASSP)*, pages 3156–3160, 2016.
- [10] D. Malioutov, M.  etin, and A. S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.*, 53(8):3010–3022, 2005.
- [11] A. Xenaki, P. Gerstoft, and K. Mosegaard. Compressive beamforming. *J. Acoust. Soc. Am.*, 136(1):260–271, 2014.
- [12] 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.
- [13] P. Gerstoft, A. Xenaki, and C.F. Mecklenbr uker. Multiple and single snapshot compressive beamforming. *J. Acoust. Soc. Am.*, 138(4):2003–2014, 2015.
- [14] P. Gerstoft, C. F. Mecklenbr uker, A. Xenaki, and S. Nannuru. Multisnapshot sparse Bayesian learning for DOA. *IEEE Signal Proc. Lett.*, 23(10):1469–1473, 2016.
- [15] S. Nannuru, K.L Gemba, P. Gerstoft, W.S. Hodgkiss, and C.F. Mecklenbr uker. Sparse Bayesian learning with uncertainty models and multiple dictionaries. *arXiv:1704.00436*, 2017.
- [16] P. Stoica and P. Babu. SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation. *Signal Proc.*, 92(7):1580–1590, 2012.
- [17] D. P. Wipf and S. Nagarajan. Beamforming using the relevance vector machine. In *Proc. 24th Int. Conf. Machine Learning*, New York, NY, USA, 2007.
- [18] D. P. Wipf and B.D. Rao. Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Proc.*, 52(8):2153–2164, 2004.
- [19] Z. Zhang and B. D Rao. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE J Sel. Topics Signal Proc.*, 5(5):912–926, 2011.
- [20] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou. An efficient maximum likelihood method for direction-of-arrival estimation via sparse Bayesian learning. *IEEE Trans. Wireless Comm.*, 11(10):1–11, Oct. 2012.
- [21] 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.
- [22] R. Giri and B. Rao. Type I and type II Bayesian methods for sparse signal recovery using scale mixtures. *IEEE Trans Signal Proc. Signal Proc.*, 64(13):3418–3428, July 2016.
- [23] J.F. B hme. Source-parameter estimation by approximate maximum likelihood and nonlinear regression. *IEEE J. Oceanic Eng.*, 10(3):206–212, 1985.
- [24] A.G. Jaffer. Maximum likelihood direction finding of stochastic sources: A separable solution. In *IEEE Int. Conf. on Acoust., Speech, and Sig. Proc. (ICASSP-88)*, volume 5, pages 2893–2896, 1988.
- [25] P. Stoica and A. Nehorai. On the concentrated stochastic likelihood function in array processing. *Circuits Syst. Signal Proc.*, 14(5):669–674, 1995.
- [26] P. Gerstoft, S. Nannuru, C. F. Mecklenbr uker, and G. Leus. Sparse Bayesian learning for doa estimation in heteroscedastic noise. *arXiv:1711.03847*, 2018.
- [27] Johann F B hme. Estimation of spectral parameters of correlated signals in wavefields. *Signal Proc.*, 11(4):329–337, 1986.
- [28] H. Krim and M. Viberg. Two decades of array signal processing research: the parametric approach. *IEEE Signal Proc. Mag.*, 13(4):67–94, 1996.
- [29] P. Stoica, B. Ottersten, M. Viberg, and R.L. Moses. Maximum likelihood array processing for stochastic coherent sources. *IEEE Trans. Signal Proc.*, 44(1):96–105, 1996.
- [30] R.L. Burden, J.D. Faires, and A.M. Burden. *Numerical analysis*. Cengage Learning, 2016. (chapter 2.2).
- [31] M. E. Tipping. Sparse Bayesian learning and the relevance vector machine. *J. Machine Learning Research*, 1:211–244, 2001.
- [32] N. Riahi and P. Gerstoft. The seismic traffic footprint: Tracking trains, aircraft, and cars seismically. *Geophys. Res. Lett.*, 42(8):2674–2681, 2015.