DOA M-ESTIMATION USING SPARSE BAYESIAN LEARNING

Christoph F. Mecklenbräuker[⋆] *Peter Gerstoft*[†] *Esa Ollila*[‡]

[⋆] Inst. of Telecommunications, TU Wien, Vienna, Austria
[†] NoiseLab, UCSD, San Diego (CA), USA
[‡] Dept. of Signal Processing and Acoustics, Aalto University, Aalto, Finland

ABSTRACT

Recent investigations indicate that Sparse Bayesian Learning (SBL) is lacking in robustness. We derive a robust and sparse Direction of Arrival (DOA) estimation framework based on the assumption that the array data has a centered (zero-mean) complex elliptically symmetric (ES) distribution with finite second-order moments. In the derivation, the loss function can be quite general. We consider three specific choices: the ML-loss for the circularly symmetric complex Gaussian distribution, the ML-loss for the complex multivariate t-distribution (MVT) with ν degrees of freedom, and the loss for Huber's M-estimator. For Gaussian loss, the method reduces to the classic SBL method. The root mean square DOA performance of the derived estimators is discussed for Gaussian, MVT, and ϵ-contaminated noise. The robust SBL estimators perform well for all cases and nearly identical with classical SBL for Gaussian noise.

Index Terms**—** DOA estimation, robust statistics, outliers, sparsity, Bayesian learning

1. INTRODUCTION

Array processing in heavy-tailed noise arises e.g. due to interference in wireless channels and clutter in radar [1, 2]. Such noise requires statistically robust array processing. Notions of statistical robustness have been defined [3, 4, 5, 6]. Qualitative robustness can be assessed by evaluating the Empirical Influence Function (EIF) [6]. The SBL approach is flexible through the usage of various priors, although Gaussian are most common [7]. Recent investigations of the EIF indicate that Sparse Bayesian Learning (SBL) is lacking in qualitative robustness [8, 9]. Previously, we developed direction of arrival (DOA) estimators for plane waves observed by a sensor array based on a complex multivariate Student t-distribution data model. A qualitatively robust and sparse DOA estimate is derived as maximum likelihood estimate based on this model. [8, 9], [6, Sec. 5.4.2]. A Bayes-optimal algorithm was proposed to estimate DOAs in the presence of impulsive noise from the perspective of SBL in [10].

We incorporate priors with potentially strong outliers by allowing various loss functions in the maximum likelihood formulation. This leads to a robust and sparse DOA estimator which is based on the assumption that the array data observations follow a centered (zero-mean) complex elliptically symmetric (ES) distribution with finite second-order moments.

In applications, the data is often normalized to constant magnitude as this is a simple form of mitigation against power variations in the array data and renders the processing robust against outliers. The normalized data only contains information through the phase of the array data. Examples are acoustics [11], source deconvolution in ocean acoustics [12, 13], and speaker localization [14, 15, 16, 17].

2. ARRAY DATA MODEL

Narrowband waves are observed on N sensors for L snapshots and the array data is $\boldsymbol{Y} = [\boldsymbol{y}_1 \dots \boldsymbol{y}_L] \in \mathbb{C}^{N\times L}$. The unknown zero-mean complex source amplitudes are the elements of $\boldsymbol{X} \in \mathbb{C}^{M\times L}$ where M is the considered number of hypothetical DOAs on a given grid. The source amplitudes are independent across sources and snapshots, i.e. x_{ml} and $x_{m'l'}$ are independent. A linear regression model relates the array data $\boldsymbol{Y}$ to the source amplitudes $\boldsymbol{X}$,

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

The M columns of the dictionary $\boldsymbol{A} \in \mathbb{C}^{N\times M}$ are the replica vectors for all hypothetical DOAs. The noise $\boldsymbol{N} \in \mathbb{C}^{N\times L}$ is assumed independent identically distributed (iid) across sensors and snapshots, zero-mean, with finite variance σ^2.

If K sources are present, the lth column of $\boldsymbol{X}$ is K-sparse and we assume that the sparsity pattern is the same for all snapshots. The active set $\mathcal{M} = \{m \in \mathbb{N} | x_{ml} \neq 0\}$ is defined and $\boldsymbol{A}_{\mathcal{M}} \in \mathbb{C}^{N\times K}$ contains the K "active" replicas from $\boldsymbol{A}$. The model (1) is underdetermined and $K < N \ll M$. The array covariance matrix takes the form

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_N, \tag{2}$$

$$\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}) \tag{3}$$

where $\boldsymbol{\gamma} = [\gamma_1 \dots \gamma_M]^T$ is the K-sparse vector of unknown source powers.

3. ROBUST AND SPARSE BAYESIAN LEARNING

SBL is derived under a joint complex multivariate Gaussian assumption on $\boldsymbol{X}$ and $\boldsymbol{N}$. Direction of arrival (DOA) estimation for plane waves using Sparse Bayesian learning (SBL) is proposed in Ref. [18, Table I]. SBL provides DOA estimates based on the sample covariance matrix

$$\boldsymbol{S}_{\boldsymbol{Y}} = \boldsymbol{Y}\boldsymbol{Y}^H/L, \tag{4}$$

of the array data sample $\boldsymbol{Y}$, where $(\cdot)^H$ denotes Hermitian transpose. The sample covariance matrix is a sufficent statistic under the jointly Gaussian assumption, but it is not robust against deviations from this assumption [8]. A numerically efficient SBL implementation is available on GitHub [19].

In the following, we derive robust and sparse Bayesian learning which can be understood as introducing a data-dependent diagonal weighting into (4),

$$\boldsymbol{R}_{\boldsymbol{Y}} = \boldsymbol{Y}\boldsymbol{D}\boldsymbol{Y}^H/L. \tag{5}$$

Different choices of $\boldsymbol{D}$ lead to different DOA estimators, e.g., $\boldsymbol{D} = \boldsymbol{I}_N$ reduces (5) to (4). Another choice leads to (20) below.

Phase-only processing of the array data is an often used adhoc approach (e.g. in seismic applications[20]) leading to robust DOA estimators [21, 22]. This is implemented by pre-processing the array data

$$\tilde{\boldsymbol{Y}} = \boldsymbol{Y} \oslash |\boldsymbol{Y}| \tag{6}$$

where $\oslash$ denotes element-wise division and $|\boldsymbol{Y}|$ is the matrix of element-wise magnitudes of $\boldsymbol{Y}$. Many DOA estimators depend on the array data solely through (4) and qualitatively robust DOA estimators can be formulated using the phase-only sample covariance matrix [22]

$$\tilde{\boldsymbol{S}}_{\boldsymbol{Y}} = \tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^H/L, \tag{7}$$

instead of (4). We note that (7) cannot be written in the form (5).

3.1. M-estimation based on ES distribution

A qualitatively robust DOA estimate was derived as maximum likelihood estimate (MLE) based on the complex multivariate t_ν-distribution as model for array data [6, Sec. 5.4.2], [9]. Assuming $\boldsymbol{X}$ and $\boldsymbol{N}$ are zero-mean, it follows from (1) that the array data $\boldsymbol{Y}$ are zero-mean.

Here, we follow a general approach based on loss functions and assume that the array data $\boldsymbol{Y}$ has a centered (zero-mean) complex elliptically symmetric (ES) distribution with positive definite Hermitian $N \times N$ covariance matrix parameter $\boldsymbol{\Sigma}$ [23, 24]. Thus

$$p(\boldsymbol{Y}|\boldsymbol{0}, \boldsymbol{\Sigma}) = \prod_{\ell=1}^{L} \det(\boldsymbol{\Sigma}^{-1}) g(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell). \tag{8}$$

An M-estimator of the covariance matrix parameter $\boldsymbol{\Sigma}$ is defined as a positive definite Hermitian $N \times N$ matrix that minimizes [6, (4.20],

$$\mathcal{L}(\boldsymbol{\Sigma}) = \frac{1}{Lb} \sum_{\ell=1}^{L} \rho(\boldsymbol{y}_\ell^H \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell) - \log\det(\boldsymbol{\Sigma}^{-1}), \tag{9}$$

where $\boldsymbol{y}_\ell$ is the ℓth array snapshot and $\rho : \mathbb{R}_0^+ \to \mathbb{R}^+$, is called the loss function. The loss function is any continuous, non-decreasing function which satisfies that $\rho(e^x)$ is convex in $-\infty < x < \infty$, cf. [6, Sec. 4.3]. The term b is a consistency factor defined as

$$b = \mathsf{E}[\psi(\|\boldsymbol{y}\|^2)]/N, \quad \boldsymbol{y} \sim \mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \boldsymbol{I}), \tag{10}$$

$$= \frac{1}{N} \int_0^\infty \psi(t/2) f_{\chi^2_{2N}}(t) \mathrm{d}t \tag{11}$$

where $\psi(t) = t\rho'(t)$ and $f_{\chi^2_{2N}}(t)$ denotes the pdf of chi-squared distribution with $2N$ degrees of freedom. To arrive from (10) to (11) we used that $\|\boldsymbol{y}\|^2 \sim (1/2)\chi^2_{2N}$. The consistency factor is inserted into (9) to guarantee that the minimizer $\hat{\boldsymbol{\Sigma}}$ of (9) is a consistent estimator of the covariance matrix $\boldsymbol{\Sigma}$ for Gaussian array data, $\boldsymbol{y}_\ell \sim \mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \boldsymbol{\Sigma})$. A derivation is given in the Appendix.

We demonstrate the loss function $\rho(\cdot)$ for three cases where they are chosen so that (9) becomes an negative log-likelihood function for the corresponding distribution: 1) **Gaussian loss** corresponds to loss function of (circular complex) Gaussian distribution:

$$\rho_{\mathrm{G}}(t) = t \tag{12}$$

in which case the objective in (9) becomes the Gaussian objective function $\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_{\boldsymbol{Y}}\} - \log\det(\boldsymbol{\Sigma}^{-1})$ and whose solution is $\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}$. In this case b in (11) becomes $b = 1$ as expected, since $\boldsymbol{S}$ is consistent to $\boldsymbol{\Sigma}$ without any scaling correction. 2) **Huber's loss** given by [6, Eq. (4.29)]

$$\rho_{\mathrm{H}}(t; c) = \begin{cases} t & \text{for } t \leqslant c^2, \\ c^2\big(\log(t/c^2) + 1\big) & \text{for } t > c^2, \end{cases} \tag{13}$$

The c^2 parameter in (13) is mapped to a qth quantile of $(1/2)\chi^2_{2N}$-distribution and we regard $q \in (0, 1)$ as a loss parameter which can be chosen by design. As default value we use $q = 0.9$. It is easy to verify that b in (11) for Huber's loss function is [6, Sec. 4.4.2],

$$b = F_{\chi^2_{2(N+1)}}(2c^2) + c^2(1 - F_{\chi^2_{2N}}(2c^2))/N, \tag{14}$$

where $F_{\chi^2_{2N}}(x)$ denotes the cdf of χ^2_{2N} distribution. 3) **MVT loss** which corresponds to loss function of (circular complex) multivariate t (MVT) distribution with ν degrees of freedom [6, Eq.(4.28)],

$$\rho_{\mathrm{T}}(t; \nu) = \frac{\nu + 2N}{2} \log(\nu + 2t). \tag{15}$$

The ν parameter in (15) is viewed as a loss parameter which can be chosen by design. In this case, the consistency factor b is not as easy to derive in closed form, but is computable using numerical integration.

Due to consistency factor used in (1), we can conclude that under nominal Gaussian data assumption, it holds that the M-estimator is consistent to covariance matrix $\boldsymbol{\Sigma}$.

3.1.1. Source Power Estimation

Similarly to Ref. [18, Sec. III.D], we regard (9) as a function of $\boldsymbol{\gamma}$ and σ^2 and compute the first order derivative

$$\frac{\partial \mathcal{L}}{\partial \gamma_m} = -\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m + \frac{1}{Lb} \sum_{\ell=1}^{L} \|\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell\|_2^2 u(\boldsymbol{y}_\ell^H \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell) \tag{16}$$

where $u(t) = \rho'(t)$ is the weight function associated with the loss function ρ. For Gaussian loss $u_{\mathrm{G}}(t) = 1$ while for MVT-loss (15) the corresponding weight function is

$$u_{\mathrm{T}}(t; \nu) = \frac{\nu + 2N}{\nu + 2t} \tag{17}$$

while for Huber's loss (13) it becomes

$$u_{\mathrm{H}}(t; c) = \begin{cases} 1, & \text{for } t \leqslant c^2 \\ c^2/t, & \text{for } t > c^2 \end{cases}. \tag{18}$$

Thus, an observation $\boldsymbol{y}_\ell$ with squared Mahalanobis distance (MD) $\boldsymbol{y}_\ell^H \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell$ smaller than c^2 receives constant weight, while observations with a larger MD are heavily down-weighted. Note the similarity of (16) with Ref. [18, Eq.(21)]. Setting (16) to zero gives

$$\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m = \boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{R}_{\boldsymbol{Y}} \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m, \tag{19}$$

where $\boldsymbol{R}_{\boldsymbol{Y}}$ is the weighted sample covariance matrix,

$$\boldsymbol{R}_{\boldsymbol{Y}} = \frac{1}{Lb} \sum_{\ell=1}^{L} u(\boldsymbol{y}_\ell^H \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell; \cdot) \boldsymbol{y}_\ell \boldsymbol{y}_\ell^H. \tag{20}$$

Equation (20) can be formulated in the form of (5) for $\boldsymbol{D} = \mathrm{diag}(u_1, \ldots, u_L)/b$ with $u_\ell = u(\boldsymbol{y}_\ell^H \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell; \cdot)$. Note that $\boldsymbol{R}_{\boldsymbol{Y}}$

input $\boldsymbol{Y} \in \mathbb{C}^{N \times L}$ array data to be analyzed
choose the desired weight function $u(\cdot;\cdot)$ and loss parameter
constant $\boldsymbol{A} \in \mathbb{C}^{N \times M}$ dictionary matrix
constants $\nu, K, j_{\max} = 1200$
initialize $\hat{\sigma}^2, \boldsymbol{\gamma}^{\text{new}}$ using (24), $\delta_{\min} = 10^{-3}$, $j = 0$
repeat
    $j = j + 1, \boldsymbol{\gamma}^{\text{old}} = \boldsymbol{\gamma}^{\text{new}}, \boldsymbol{\Gamma} = \operatorname{diag}(\boldsymbol{\gamma}^{\text{new}})$
    $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^H + \hat{\sigma}^2 \boldsymbol{I}_N$ (2)
    $\boldsymbol{R}_{\boldsymbol{Y}} = \frac{1}{Lb} \sum_{\ell=1}^{L} u(\boldsymbol{y}_\ell^H \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell; \cdot) \boldsymbol{y}_\ell \boldsymbol{y}_\ell^H$ (20)
    $\gamma_m^{\text{new}} = \gamma_m^{\text{old}} \left(\frac{\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{R}_{\boldsymbol{Y}} \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m}{\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m} \right)$ (21)
    $\mathcal{M} = \{m \in \mathbb{N} \mid K$ largest peaks in $\boldsymbol{\gamma}^{\text{new}}\}$ active set
    $\boldsymbol{A}_{\mathcal{M}} = [\boldsymbol{a}_{m_1}, \ldots, \boldsymbol{a}_{m_K}]$
    $\hat{\sigma}^2 = \hat{\sigma}_R^2 = \frac{\operatorname{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}} \boldsymbol{A}_{\mathcal{M}}^+) \boldsymbol{R}_{\boldsymbol{Y}}\right]}{N-K}$ (23)
    $\delta = \frac{\|\boldsymbol{\gamma}^{\text{new}} - \boldsymbol{\gamma}^{\text{old}}\|_1}{\|\boldsymbol{\gamma}^{\text{old}}\|_1}$
until $\delta \leq \delta_{\min}$ or $j > j_{\max}$
output $\mathcal{M}, \boldsymbol{\gamma}^{\text{new}}, \hat{\sigma}^2$

Table 1. DOA M-Estimation using Sparse Bayesian Learning

can be understood as an adaptively weighted sample covariance matrix [6, Sec. 4.3]. $\boldsymbol{R}_{\boldsymbol{Y}}$ is Fisher consistent for the covariance matrix when $\boldsymbol{Y}$ follows a Gaussian, i.e. $\mathsf{E}[\boldsymbol{R}_{\boldsymbol{Y}}] = \boldsymbol{\Sigma}$ thanks to the consistency factor b [6, Sec. 4.4.1]. We multiply (19) by γ_m and obtain the fixed-point equation

$$\begin{aligned} \gamma_m &= \gamma_m \frac{\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{R}_{\boldsymbol{Y}} \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m}{\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m} \quad \forall m \in \{1, \ldots, M\}, \\ &= \gamma_m \frac{\frac{1}{Lb} \sum_{\ell=1}^{L} \left| \boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell \sqrt{u(\boldsymbol{y}_\ell^H \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell)} \right|^2}{\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m} \end{aligned} \tag{21}$$

which is the basis for an iteration to solve for $\boldsymbol{\gamma}$ numerically. The active set $\mathcal{M}$ is then selected as either the K largest entries of $\boldsymbol{\gamma}$ or the entries with γ_m exceeding a threshold.

3.1.2. Noise Variance Estimation

The original SBL algorithm exploits Jaffer's necessary condition [25, Eq. (6)] which leads to the noise subspace based estimate [26, Eq. (15)], [18, Sec. III.E],

$$\hat{\sigma}_S^2 = \frac{\operatorname{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}} \boldsymbol{A}_{\mathcal{M}}^+) \boldsymbol{S}_{\boldsymbol{Y}}\right]}{N-K}, \tag{22}$$

where $(\cdot)^+$ denotes the Moore-Penrose pseudo inverse. This noise variance estimate works well with DOA estimation [18, 27, 28] without outliers in the array data. For ES distributed array data, we estimate the noise based on (19). This results in the robust noise variance estimate

$$\hat{\sigma}_R^2 = \frac{\operatorname{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}} \boldsymbol{A}_{\mathcal{M}}^+) \boldsymbol{R}_{\boldsymbol{Y}}\right]}{N-K}, \tag{23}$$

for the full derivation see [9]. For Gaussian loss function, $\boldsymbol{R}_{\boldsymbol{Y}} = \boldsymbol{S}_{\boldsymbol{Y}}$, the expressions (22), (23) are identical.

3.1.3. Initialization

In our algorithm we need to give initial values of source signal powers $\boldsymbol{\gamma}$ and the noise variance σ^2. These are computed as follows:

$$\begin{cases} \gamma_m^{\text{new}} = \dfrac{\boldsymbol{a}_m^H \mathbf{S}_Y \boldsymbol{a}_m}{\|\boldsymbol{a}_m\|^4}, \; m = 1, \ldots, M \\ \mathcal{M} = \{m \in \mathbb{N} \mid K \text{ largest peaks in } \boldsymbol{\gamma}^{\text{new}}\} \\ \hat{\sigma}^2 = \hat{\sigma}_S^2 = \dfrac{\operatorname{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}} \boldsymbol{A}_{\mathcal{M}}^+) \boldsymbol{S}_{\boldsymbol{Y}}\right]}{N-K} \end{cases} \tag{24}$$

4. SIMULATION RESULTS

DOA estimation performance is assessed by DOA root mean squared error (RMSE) versus array signal to noise ratio (ASNR) by numerical simulations using synthetic array data.

Array data $\boldsymbol{Y}$ are generated for scenarios with a single plane wave ($K = 1$) with complex circularly symmetric zero-meaan Gaussian amplitude from DOA $-45°$ and additive noise observed by a uniform linear array with $N = 20$ elements at half-wavelength spacing. Here, $\text{ASNR} = N\gamma_1/\sigma^2$, cf. [29, Eq. (8.112)]. The dictionary $\boldsymbol{A}$ contains $M = 18001$ replica vectors which are computed for the high resolution DOA grid $\{0.00°, 0.01°, \ldots, 179.99°\}$. Three types of noise $\boldsymbol{N} = [\boldsymbol{n}_1 \ldots \boldsymbol{n}_L]$ in (1) are simulated:
Gaussian: $\boldsymbol{n}_\ell \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$, the nominal assumption.
MVT: We first generate independent $\boldsymbol{z}_\ell =\sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ using (2) and $s \sim \chi_\nu^2$ then $\boldsymbol{y}_\ell = \boldsymbol{z}_\ell \sqrt{\nu/s} \sim \mathbb{C}t_\nu$-distributed, cf. [8] and [6, Sec. 4.2.2]. The limiting distribution of $\mathbb{C}t_\nu$-distributed noise for $\nu \to \infty$ is Gaussian.
ϵ-contaminated: $\boldsymbol{n}_\ell$ is drawn from $\mathbb{C}\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$ with probability $(1-\epsilon)$ and with outlier probability ϵ from $\mathbb{C}\mathcal{N}(\boldsymbol{0}, (\lambda\sigma)^2 \boldsymbol{I}_N)$, where λ is called the outlier strength. The resulting noise covariance matrix is $(1 - \epsilon + \epsilon\lambda^2)\sigma^2 \boldsymbol{I}_N$. The limiting distribution of ϵ-contaminated noise for $\epsilon \to 0$ and any constant $\lambda > 0$ is Gaussian.

The proposed DOA M-estimation algorithm using SBL is displayed in Table 1. For performance comparison, we evaluate DOA estimates by the conventional beamformer (CBF), by the SBL implementation `SBL_v4.m` [19] and by SBL applied to phase-only data (6) for identical synthetic data realizations $\boldsymbol{Y}$. Additionally, the Cramér-Rao Bound (CRB) for DOA estimation for a single source in additive white Gaussian noise (AWGN) is shown [29, Eqs. (8.130) with (8.702)]. A semiparametric stochastic CRB for DOA estimation under the (complex) ES data model assumption is in [30].

Figure 1 shows results for RMSE of DOA estimates in scenarios with $L = 25$ snapshots and $N = 20$ sensors. RMSE is averaged over $4 \cdot 10^4$ iid realizations of DOA estimates from array data $\boldsymbol{Y}$. There are more snapshots L than sensors N, ensuring full rank $\boldsymbol{S}_{\boldsymbol{Y}}$, $\tilde{\boldsymbol{S}}_{\tilde{\boldsymbol{Y}}}$, and $\boldsymbol{R}_{\boldsymbol{Y}}$ almost surely.

Simulations for Gaussian noise are shown in Fig. 1(a). Here, the CBF is the maximum-likelihood DOA estimator and approaches the Gaussian CRB for ASNR greater 4 dB. All four SBL-type DOA estimators perform slightly worse than the CBF. We see that the SBL4 using MVT-loss (label "SBL4-T" in the legend) for loss parameter $\nu = 2.1$ and using Huber's-loss (label "SBL4-H") for loss parameter $q = 0.9$ perform equally due to the consistency factor b introduced in (13) and (15). SBL4 with and without phase-only pre-processing (labels "SBL4-phase" and "SBL4-G", respectively) perform very similarly and slightly worse than the others.

Figure 1(b) shows simulations for MVT noise with noise parameter $\nu = 2.1$ being small. We observe that SBL-T performs best,

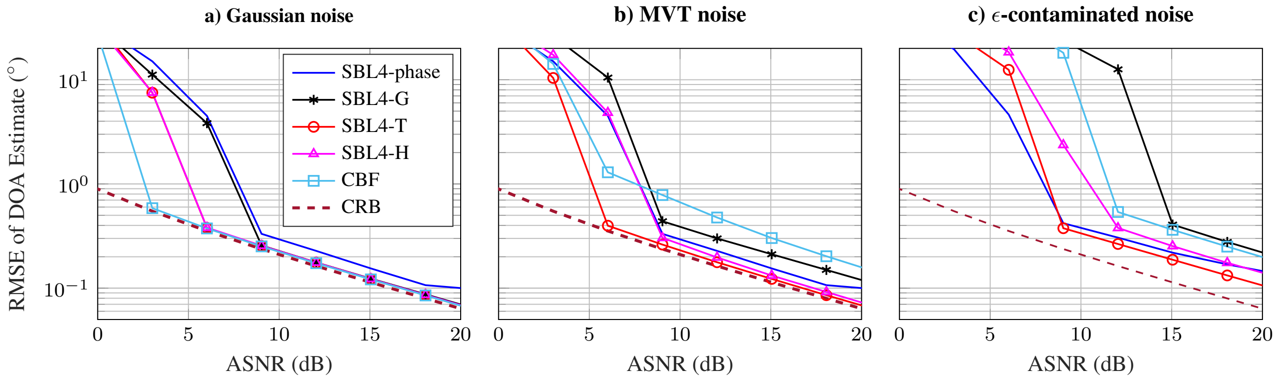

Fig. 1. RMSE of DOA estimators vs. ASNR. Simulation for uniform line array, $N = 20$ sensors, $L = 25$ array snapshots, and dictionary size $M = 18001$ corresponding to DOA resolution 0.01°, averaged over $10^4/L$ realizations. Noise: (a) Gaussian, (b) MVT ($\nu = 2.1$), (c) ϵ-contaminated ($\epsilon = 0.05$, $\lambda = 10$). The CRB line is for Gaussian noise model.

closely followed by phase-only processing with SBL4-G and SBL-H which perform equally.

Here, the loss parameter ν used by SBL4-T is identical to the MVT noise parameter ν and thus SBL4-T is expected to work well in this case. In Fig. 1(b), SBL4-T closely follows the Gaussian CRB for $\mathrm{ASNR} > 6\,\mathrm{dB}$, although a small gap at high ASNR remains. Instead of comparing to the Gaussian CRB, a semiparametric stochastic CRB for DOA estimation under the ES data model would be more appropriate for MVT noise [31, 30]. The assumption that ν is known *a priori* is somewhat unrealistic. Original SBL4 and CBF exhibit largest RMSE at high ASNR.

Results for ϵ-contaminated noise are shown in Fig. 1(c) for outlier probability $\epsilon = 0.05$ and outlier strength $\lambda = 10$. Phase-only processing with SBL4 shows lowest RMSE below ASNR$=$ 7 dB, followed by the SBL4-T using loss parameter $\nu = 2.1$ and with slight penalty Huber's loss. Poorest RMSE exhibit CBF and SBL4-G indicating strong impact of outliers. The asymptotic gap to the Gaussian CRB that is visible in Fig. 1(c) is evaluated in [32, Sec. III].

5. CONCLUSION

A robust and sparse DOA estimate was derived based on array data following a zero-mean complex elliptically symmetric distribution with finite second-order moments. The framework is based on a loss functions which can be chosen freely subject to certain existence and uniqueness conditions. Three choices for loss function are discussed: the ML-loss function for the circular complex multivariate t-distribution with ν degrees of freedom and the loss function for Huber's M-estimator. For Gaussian loss, the method reduces to Sparce Bayesian Learning.

Finally, we discussed and compare the root mean square DOA estimator performance of the derived estimators for Gaussian, MVT, and ϵ-contaminated noise. The robust SBL estimators perform well in simulations for MVT and ϵ-contaminated noise and nearly identical with classical SBL for Gaussian noise.

6. APPENDIX

Here the consistency factor b is evaluated for Huber's loss function (13) and the MVT loss function (15). It is well known that for elliptical distributions an M-estimator is consistent estimator of $\sigma\mathbf{\Sigma}$, where the constant σ is a solution to [23, eq. (49)]:

$$1 = \mathsf{E}\big[\psi(\boldsymbol{y}^H\mathbf{\Sigma}^{-1}\boldsymbol{y}/\sigma)\big]/N, \tag{25}$$

where $\psi(t) = tu(t) = t\rho'(t)$ as defined earlier. Assuming $\boldsymbol{y} \sim \mathbb{C}\mathcal{N}_N(\mathbf{0}, \mathbf{\Sigma})$, then we can scale the chosen loss function $\rho(t)$ such that (25) holds for $\sigma = 1$. Namely, for

$$\rho_b(t) = \rho(t)/b \qquad (\text{ and } \quad u_b(t) = u(t)/b\,),$$

where b is a scaling constant defined in (10), it clearly holds that $1 = \mathsf{E}\big[\psi_b(\boldsymbol{y}^H\mathbf{\Sigma}^{-1}\boldsymbol{y})\big]/N$ for $\psi_b(t) = tu_b(t)$. This implies that $\sigma = 1$ and that the M-estimator with loss $\rho_b(\cdot)$ is consistent to the covariance matrix $\mathbf{\Sigma}$ when data is from $\mathbb{C}\mathcal{N}_N(\mathbf{0}, \mathbf{\Sigma})$ distribution.

For Huber's loss function (13) the b in (11) can be solved in closed form as

$$b = \frac{1}{N}\int_0^\infty (t/2)u_{\mathrm{H}}(t/2)f_{\chi^2_{2N}}(t)\mathrm{d}t \tag{26}$$

$$= \frac{1}{2N}\int_0^{2c^2} tf_{\chi^2_{2N}}(t)\mathrm{d}t + \frac{1}{N}\int_{2c^2}^\infty c^2 f_{\chi^2_{2N}}(t)\mathrm{d}t \tag{27}$$

$$= F_{\chi^2_{2(N+1)}}(2c^2) + c^2(1 - F_{\chi^2_{2N}}(2c^2))/N, \tag{28}$$

where $F_{\chi^2_n}(x) = \mathrm{P}\{X \leq x\}$ is the cumulative distribution of a central χ^2 distributed random variable X with n degrees of freedom. For the MVT loss in (15) we evaluate (11) by numerical integration.

7. REFERENCES

[1] L. Clavier, T. Pedersen, I. Larrad, M. Lauridsen, and M. Egan. Experimental evidence for heavy tailed interference in the IoT. *IEEE Communications Letters*, 25(3):692–695, 2021.

[2] F. Gini and M. Greco. Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter. *Signal Processing*, 82(12):1847–1859, 2002.

[3] P. J. Huber. *Robust statistics*. Springer, 2011.

[4] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics: the approach based on influence functions*, volume 196. John Wiley & Sons, 2011.

[5] A.M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma. Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. *IEEE Signal Processing Magazine*, 29(4):61–80, 2012.

[6] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma. *Robust Statistics for Signal Processing*. Cambridge University Press, 2018.

[7] P. Gerstoft and P. D. Bromirski. “Weather bomb” induced seismic signals. *Science*, 353(6302):869–870, 2016.

[8] E. Ollila and V. Koivunen. Robust antenna array processing using M-estimators of pseudo-covariance. In *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, (PIMRC 2003)*, volume 3, pages 2659–2663, 2003.

[9] C. F. Mecklenbräuker, P. Gerstoft, and E. Ollila. Qualitatively robust Bayesian learning for DOA from array data using M-estimation of the scatter matrix. In *IEEE/ITG Workshop on Smart Antennas*, Sophia-Antipolis, France, Nov. 2021.

[10] J. Dai and H. C. So. Sparse Bayesian learning approach for outlier-resistant direction-of-arrival estimation. *IEEE Transactions on Signal Processing*, 66(3):744–756, 2018.

[11] P. Gerstoft, W. S. Hodgkiss, M. Siderius, C.-F. Huang, and C.H. Harrison. Passive fathometer processing. *J. Acoust. Soc. Am.*, 123(3):1297–1305, 2008.

[12] K.G. Sabra and D.R. Dowling. Blind deconvolution in ocean waveguides using artificial time reversal. *J. Acoust. Soc. Am.*, 116(1):262–271, 2004.

[13] K.G. Sabra, H-C. Song, and D.R. Dowling. Ray-based blind deconvolution in ocean sound channels. *J. Acoust. Soc. Am.*, 127(2):EL42–EL47, 2010.

[14] S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Transactions on Signal Processing*, 49(8):1614–1626, 2001.

[15] M. J. Bianco, S. Gannot, E. Fernandez-Grande, and P. Gerstoft. Semi-supervised source localization in reverberant environments with deep generative modeling. *IEEE Access*, 9:84956–84970, 2021.

[16] O. Schwartz and S. Gannot. Speaker tracking using recursive em algorithms. *IEEE Trans. Audio, Speech, Language Process.*, 22(2):392–402, 2014.

[17] Y. Dorfan and S. Gannot. Tree-based recursive expectation-maximization algorithm for localization of acoustic sources. *IEEE Trans. Audio, Speech and Lang. Process.*, 23(10):1692–1703, 2015.

[18] P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, and S. Nannuru. Multisnapshot sparse Bayesian learning for DOA. *IEEE Signal Process. Lett.*, 23(10):1469–1473, 2016.

[19] Y. Park, S. Nannuru, K. Gemba, and P. Gerstoft. SBL4 from NoiseLab. In `https://github.com/gerstoft/SBL`. GitHub, 2020.

[20] P. Gerstoft, M. C. Fehler, and K. G. Sabra. When Katrina hit California. *Geophys. Res. Lett.*, 33(17):L17308, 2006.

[21] H. Groll, E. Zöchmann, P. Gerstoft, and C. F. Mecklenbräuker. Global robustness characterization of DOA estimators by breakdown detection. In *IEEE CAMSAP*, Guadeloupe, French West Indies, Dec. 2019.

[22] C. F. Mecklenbräuker, P. Gerstoft, E. Zöchmann, and H. Groll. Robust estimation of DOA from array data at low SNR. *Signal Processing*, 166:107262, 2020.

[23] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor. Complex elliptically symmetric distributions: survey, new results and applications. *IEEE Transactions on Signal Processing*, 60(11):5597–5625, 2012.

[24] M. Mahot, F. Pascal, P. Forster, and J.-P. Ovarlez. Asymptotic properties of robust complex covariance matrix estimates. *IEEE Transactions on Signal Processing*, 61(13):3348–3356, 2013.

[25] A. G. 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, pages 2893–2896, 1988.

[26] 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.

[27] S. Nannuru, A. Koochakzadeh, K.L Gemba, P. Pal, and P. Gerstoft. Sparse bayesian learning for beamforming using sparse linear arrays. *J. Acoust. Soc. Am.*, 144(5):2719–2729, 2018.

[28] Y. Park, F. Meyer, and P. Gerstoft. Sequential sparse bayesian learning for time-varying direction of arrival. *J. Acoust. Soc. Am.*, 149(3):2089–2099, 2021.

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

[30] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, and M. Rangaswamy. Semiparametric CRB and Slepian–Bangs formulas for complex elliptically symmetric distributions. *IEEE Transactions on Signal Processing*, 67(20):5352–5364, Oct. 2019.

[31] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir, , and M. Rangaswamy. Semiparametric inference and lower bounds for real elliptically symmetric distributions. *IEEE Transactions on Signal Processing*, 67(1):164–177, Jan. 2019.

[32] R.J. Kozick and B.M. Sadler. Maximum-likelihood array processing in non-gaussian noise with gaussian mixtures. *IEEE Transactions on Signal Processing*, 48(12):3520–3535, 2000.