ROBUST SPARSE BAYESIAN LEARNING FOR DOA

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

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

ABSTRACT

We formulate statistically robust Sparse Bayesian Learning (SBL) for Direction of Arrival (DOA) estimation from Complex Elliptically Symmetric (CES) data using a general approach based on loss functions. Simulation results for DOA estimation are obtained for several choices of loss functions: Gauss, multivariate t (MVT), Huber, and Tyler. The root mean square DOA error is discussed for Gaussian, MVT, and ϵ-contaminated data. The robust SBL estimators perform well in the presence of outliers and for heavy-tailed data and almost like classical SBL for Gaussian data.

1. INTRODUCTION

Heavy-tailed data arises e.g. due to clutter in radar [1] and interference in wireless links [2]. Such data demand a statistically robust approach [3].

Sparse Bayesian Learning (SBL) was originally derived under a multivariate Gaussian assumption on the data [4]. Direction of arrival (DOA) estimation for plane waves using SBL is proposed in Ref. [5, Table I]. The original SBL is lacking in statistical robustness [6, 7], but the SBL approach can incorporate various priors. A Bayes-optimal algorithm was proposed to estimate DOAs in the presence of impulsive noise from the perspective of SBL in [8]. In the following, we derive robust and sparse Bayesian learning which can be understood as introducing a data-dependent weighting into the sample covariance matrix (SCM) estimate.

First, M-estimation approach to SBL for DOA based on the Complex Elliptically Symmetric (CES) data model with unknown source and noise variances is formulated. This is similar to stochastic likelihood (type-II likelihood) maximization in array processing [9, 10].

Next, the likelihood function is derived and various priors are incorporated with potentially strong outliers. This leads to a robust and sparse DOA estimator which is based on the assumption that the data observations follow a centered (zero-mean) CES distribution with finite second-order moments.

2. COMPLEX ELLIPTICALLY SYMMETRIC DATA

Narrowband waves are observed on N sensors for L snapshots $\boldsymbol{y}_\ell$ and the data is $\boldsymbol{Y} = [\boldsymbol{y}_1 \dots \boldsymbol{y}_L] \in \mathbb{C}^{N\times L}$. The snapshots $\boldsymbol{y}_\ell$ are modeled as independent identically distributed (iid) with Complex Elliptically Symmetric (CES) distribution,

$$\boldsymbol{y}_\ell \sim \mathrm{CES}(\boldsymbol{0}, \boldsymbol{\Sigma}, g)\,, \tag{1}$$

where $\boldsymbol{0}$ is the mean, $\boldsymbol{\Sigma}$ is the scatter matrix and g the density generator. Its probability density function (pdf) is of the form

$$p_{\boldsymbol{y}}(\boldsymbol{y}_\ell) = C\ (\det \boldsymbol{\Sigma})^{-1}\, g(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell)\,. \tag{2}$$

where C is an irrelevant normalization constant. We further assume that the covariance matrix $\mathrm{cov}(\boldsymbol{y}_\ell) = \mathsf{E}(\boldsymbol{y}_\ell \boldsymbol{y}_\ell^{\mathsf{H}})$ exists and equals the scatter matrix $\boldsymbol{\Sigma}$, cf. [7]. The scatter matrix is then modeled as

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{H}} + \sigma^2 \boldsymbol{I}_N, \tag{3}$$

$$\boldsymbol{\Gamma} = \mathrm{cov}(\boldsymbol{x}_\ell) = \mathrm{diag}(\boldsymbol{\gamma}) \tag{4}$$

where $\boldsymbol{A} \in \mathbb{C}^{N\times M}$ is a dictionary matrix, $\boldsymbol{\gamma} = [\gamma_1 \dots \gamma_M]^{\mathsf{T}}$ is the K-sparse vector of unknown source powers, $M \gg N$, and σ^2 represents the noise power.

The data model (1) has been used for modelling heavy-tailed non-Gaussian data, most notably clutter in radar [11–13]. We use CES assumptions differently from [11–13]. where the model has been used to model noise, clutter, or interference. We assume that the snapshots follow a CES distribution. The data model (1) includes the additive white Gaussian noise (AWGN) model as a special case and is general enough to include data models with heavy-tails or outliers.

For numerical performance evaluations of the derived M-estimator of DOA, three data models are used in Sec. 4: Gaussian, MVT, and ϵ-contaminated.

2.1. Dictionary for plane wave arrivals

The M columns of the dictionary $\boldsymbol{A} = [\boldsymbol{a}_1 \dots \boldsymbol{a}_M]$ are the replica vectors for all hypothetical DOAs. For a uniform linear array (ULA), the dictionary matrix elements are $A_{nm} = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(n-1)d\sin\theta_m}$ where λ is the wavelength and d is the element spacing. The dictionary's DOA grid is defined as $\theta_m = -90° + (m-1)\delta,\ \forall m = 1,\dots,M$ where δ is the dictionary's angular grid resolution, $\delta = 180°/(M-1)$.

2.2. Traditional signal plus noise model

The unknown zero-mean complex source amplitudes are the elements of $\boldsymbol{X} = [\boldsymbol{x}_1 \dots \boldsymbol{x}_L] \in \mathbb{C}^{M\times L}$ where M is the considered number of hypothetical DOAs on the given grid $\{\theta_1, \dots, \theta_M\}$. The source amplitudes are independent across sources and snapshots, i.e., x_{ml} and $x_{m'l'}$ are independent for $(m,l) \neq (m',l')$.

The noise $\boldsymbol{N} = [\boldsymbol{n}_1 \dots \boldsymbol{n}_L] \in \mathbb{C}^{N\times L}$ is assumed independent identically distributed (iid) across sensors and snapshots, zero-mean, with finite variance σ^2 for all n, ℓ.

The commonly assumed Gaussian data model is recovered for $g(t) = \mathrm{e}^{-t}$, cf. Table 1, Then

$$\boldsymbol{y}_\ell = \boldsymbol{A}\boldsymbol{x}_\ell + \boldsymbol{n}_\ell \,. \tag{5}$$

The source and noise amplitudes are jointly Gaussian and independent of each other, i.e. $\boldsymbol{x}_\ell \sim \mathbb{C}\mathcal{N}_M(\boldsymbol{0}, \boldsymbol{\Gamma})$ and $\boldsymbol{n}_\ell \sim \mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_N)$. This model results in Gaussian data, $\boldsymbol{y}_\ell \sim \mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \boldsymbol{\Sigma})$. Model (5) was assumed in [5] while here the more general model (1) is used.

2.3. MVT distributed snapshots

In array processing applications, the complex Multi-Variate t-distribution (MVT distribution) [14, 15] can be used as an alternative to the Gaussian distribution in the presence of outliers because it has heavier tails than the Gaussian distribution. The MVT-distribution is a suitable choice for such data and provides a parametric approach to robust statistics [3, 6]. The complex MVT distribution is a special case of the CES distribution, cf. Table 1.

2.4. Source sparsity pattern and active set

We assume that K sources are present and the sparsity pattern is the same for all snapshots. The sparsity pattern is modeled by the active set

$$\mathcal{M} = \{m \in \{1, \dots, M\} \,|\gamma_m \neq 0\} = \{m_1, \dots, m_K\}. \tag{6}$$

In the traditional signal model, $x_{m\ell} = 0$ for all $m \notin \mathcal{M}$ and all ℓ. The ℓth column of $\boldsymbol{X}$ is K-sparse.

The K "active" replica vectors are aggregated in the active dictionary,

$$\boldsymbol{A}_\mathcal{M} = [\boldsymbol{a}_{m_1} \dots \boldsymbol{a}_{m_K}] \in \mathbb{C}^{N\times K}, \tag{7}$$

with its kth column vector $\boldsymbol{a}_{m_k}$, where $m_k \in \mathcal{M}$.

3. M-ESTIMATION BASED ON CES DISTRIBUTION

We follow a general approach based on loss functions and assume that the data $\boldsymbol{Y}$ has a CES distribution with zero mean $\boldsymbol{0}$ and positive definite Hermitian $N \times N$ covariance matrix parameter $\boldsymbol{\Sigma}$ [16, 17]. Thus

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

An M-estimator of the covariance matrix $\boldsymbol{\Sigma}$ is defined as a positive definite Hermitian $N \times N$ matrix that minimizes the objective function [3, (4.20)],

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

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. [3, Sec. 4.3]. Note that the objective function (9) is a penalized sample average of the chosen loss function ρ where the penalty term is $\log\det\boldsymbol{\Sigma}$. A specific choice of loss function ρ renders (9) equal to the negative log-likelihood of $\boldsymbol{\Sigma}$ when the data are CES distributed with density generator $g(t) = \mathrm{e}^{-\rho(t)}$ [18]. If the loss function is chosen, e.g., as $\rho(t) = t$ then (9) becomes the negative log-likelihood function for $\boldsymbol{\Sigma}$ for Gaussian data.

The term b is a fitting coefficient, called consistency factor, which renders the minimizer of the objective function (9) to be equal to $\boldsymbol{\Sigma}$ when the data are Gaussian, thus

$$b = \frac{1}{N}\int_0^\infty \psi(t/2) f_{\chi^2_{2N}}(t)\mathrm{d}t \tag{10}$$

where $\psi(t) = t\,\mathrm{d}\rho(t)/\mathrm{d}t$ and $f_{\chi^2_{2N}}(t)$ denotes the pdf of chi-squared distribution with $2N$ degrees of freedom.

Minimizing (9) with b according to (10) results in a consistent M-estimator of the covariance matrix $\boldsymbol{\Sigma}$ when the objective function is derived under a given non-Gaussian data assumption (as in Sec. 3.1) but is in fact Gaussian ($\boldsymbol{y}_\ell \sim \mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \boldsymbol{\Sigma})$).

3.1. Loss functions

Four loss functions $\rho(\cdot)$ are used in simulations. These are summarized in Table 1 together with their weight functions $u(\cdot;\cdot)$, density generators $g(\cdot)$, and consistency factors b.

3.2. M-Estimate of Source Power

Similarly to Ref. [5, Sec. III.D], we regard (9) as a function of $\boldsymbol{\gamma}$ and σ^2 and compute the first order derivative This gives

$$\boldsymbol{a}_m^\mathsf{H}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m = \boldsymbol{a}_m^\mathsf{H}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m, \tag{11}$$

where $\boldsymbol{R}_\boldsymbol{Y}$ is the adaptively weighted SCM [3, Sec. 4.3],

$$\boldsymbol{R}_\boldsymbol{Y} = \frac{1}{Lb}\sum_{\ell=1}^{L} u(\boldsymbol{y}_\ell^\mathsf{H}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell;\cdot)\boldsymbol{y}_\ell\boldsymbol{y}_\ell^\mathsf{H} \tag{12}$$

which is Fisher consistent for the covariance matrix when $\boldsymbol{Y}$ follows a Gaussian, i.e. $\mathsf{E}[\boldsymbol{R}_\boldsymbol{Y}] = \boldsymbol{\Sigma}$ due to the consistency factor b [3, Sec. 4.4.1]. We multiply (11) by γ_m and obtain the fixed-point equation

$$\gamma_m = \gamma_m \frac{\boldsymbol{a}_m^\mathsf{H}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m}{\boldsymbol{a}_m^\mathsf{H}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m} \;\forall m \in \{1, \dots, M\}, \tag{13}$$

loss name	data density generator $g(t)$	loss $\rho(t)$	weight $u(t;\cdot)$	$\psi(t)$	loss parameter	consistency factor b
	(2)	$-\log g(t)$	$\rho'(t)$	$tu(t;\cdot)$		(10)
Gauss	e^{-t}	t	1	t	n/a	1.000
MVT	$(1+t/\nu)^{-(\nu+2N)/2}$	$\frac{\nu+2N}{2}\log(\nu+2t)$	$\frac{\nu+2N}{\nu+2t}$	$\frac{\nu+2N}{2+\nu/t}$	$\nu=2.1$	0.998
Huber	$\exp(-\rho_{\mathrm{Huber}}(t;c))$	[3, Eq. (4.29)]	$=\begin{cases}1 \text{ if } t<c^2\\ c^2/t \text{ else}\end{cases}$	$t\,u_{\mathrm{Huber}}(t)$	$q=0.9$	0.987
Tyler	t^{-N}	$N\log t$	N/t	N	n/a	$1/\hat{\tau}$

Table 1. Density generators used for CES data generation, loss and weight functions used in DOA M-estimation, consistency factor b is for $N=20$ and for Tyler, $\hat{\tau}$ is a robust estimate of $\mathrm{tr}(\boldsymbol{\Sigma})/N$.

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.

3.3. M-Estimate of Noise Power

The original SBL algorithm exploits Jaffer's necessary condition [10, Eq. (6)] for the noise variance estimate. For CES data, the noise is estimated robustly [6] based on (11),

$$\hat{\sigma}_R^2 = \mathrm{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^{+})\boldsymbol{R}_{\boldsymbol{Y}}\right]/(N-K), \qquad (14)$$

where $(\cdot)^+$ denotes the Moore-Penrose pseudo inverse.

To stabilize the noise variance M-estimate (14) for non-Gauss loss, we define lower and upper bounds for $\hat{\sigma}^2$ and enforce $\sigma^2_{\text{floor}} \le \hat{\sigma}^2 \le \sigma^2_{\text{ceil}}$ by

$$\hat{\sigma}^2 = \max(\min(\hat{\sigma}_R^2, \sigma^2_{\text{ceil}}), \sigma^2_{\text{floor}}) \qquad (15)$$

Tyler's M-estimator is unique only up to a scale which affects the noise variance estimate $\hat{\sigma}_R^2$. For this reason, we normalize $\boldsymbol{R}_{\boldsymbol{Y}}$ to trace 1 to remove this ambiguity if Tyler loss is used.

3.4. Algorithm

The proposed DOA M-estimation algorithm using SBL is displayed in Table 2 with the following remarks:

DOA grid pruning: To speed up the iterations, we introduce the pruned DOA grid $\mathcal{P}$ for focusing computational resources on those DOAs which are associated with source power estimates exceeding a minimum threshold γ_{floor} .The pruned DOA grid is formally defined as an index set,

$$\mathcal{P} = \{p \in \{1,\ldots,M\} \mid \boldsymbol{\gamma}_p^{\text{new}} \ge \gamma_{\text{floor}}\} = \{p_1,\ldots,p_P\} \quad (16)$$

where $\gamma_{\text{floor}} = \gamma_{\text{range}} \max(\boldsymbol{\gamma}^{\text{new}})$ and we choose $\gamma_{\text{range}} = 10^{-3}$.

Convergence Criterion: The DOA Estimates returned by the iterative algorithm in Table 2 are obtained from the active set $\mathcal{M}$. Therefore, the active set is monitored for changes in its elements to determine whether the algorithm has converged. If $\mathcal{M}$ has not changed during the last $z \in \mathbb{N}$ iterations then the repeat-until loop (lines 14–29 in Table 2) is exited. Here z is a tuning parameter which allows to trade off computation time against DOA estimate accuracy. To ensure that the iterations always terminate, the maximum iteration count is defined as $j_{\max}$ with $z < j_{\max}$.

4. SIMULATION RESULTS

Simulations are carried out for evaluating the root mean squared error (RMSE) of DOA versus array signal to noise ratio (ASNR) based on synthetic array data $\boldsymbol{Y}$. Synthetic array data are generated for two scenarios as listed in Table 3. The source amplitudes $\boldsymbol{x}_\ell$ in (4) are complex circularly symmetric zero-mean Gaussian. The wavefield is modeled by (1) and observed by a ULA with $N = 20$ sensors at $\lambda/2$ spacing. The dictionary $\boldsymbol{A}$ consists of $M = 18001$ replica vectors and the dictionary's angular grid resolution is $\delta = 180°/(M-1) = 0.01°$, cf. Sec. 2.1.

The RMSE of the DOA estimates over $N_{\text{run}} = 250$ simulation runs with random data realizations is used for evaluating the algorithm's performance,

$$\mathrm{RMSE} = \sqrt{\sum_{r=1}^{N_{\text{run}}}\sum_{k=1}^{K}\frac{[\min(|\hat{\theta}_k^r - \theta_k^r|, e_{\max})]^2}{K\,N_{\text{run}}}}, \qquad (17)$$

where θ_k^r is the true DOA of the k source and $\hat{\theta}_k^r$ is the corresponding estimated DOA in the rth run when K sources are present in the scenario. This RMSE definition is a specialization of the optimal subpattern assignment (OSPA) when K is known, cf. [20]. We use $e_{\max} = 10°$ in (17). Thus maximum RMSE is $10°$.

The synthetic CES data for Gaussian and MVT cases are generated according to model (1) with (3) and generators listed in Table 1. We also generated ϵ-contaminated data for evaluation of RMSE. This heavy-tailed data model is not covered by model (1) with the assumptions in Sec. 2. Instead, the noise $\boldsymbol{n}$ is drawn with probability $(1-\epsilon)$ from a $\mathbb{CN}(\boldsymbol{0}, \sigma_1^2\boldsymbol{I})$ and with probability ϵ from a $\mathbb{CN}(\boldsymbol{0}, \lambda^2\sigma_1^2\boldsymbol{I})$, where λ is the outlier strength. Thus, $\boldsymbol{y}_\ell$ is drawn from $\mathbb{CN}(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{H}}+\sigma_1^2\boldsymbol{I}_N)$, using (3) with probability $(1-\epsilon)$ and with outlier probability ϵ from $\mathbb{CN}(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{H}} + (\lambda\sigma_1)^2\boldsymbol{I}_N)$. The resulting noise covariance matrix is $\sigma^2\boldsymbol{I}_N$ similar to the other models, but with

$$\sigma^2 = (1-\epsilon+\epsilon\lambda^2)\sigma_1^2. \qquad (18)$$

input $\boldsymbol{Y} \in \mathbb{C}^{N\times L}$ data to be analyzed
select the weight function $u(\cdot;\cdot)$ and loss parameter
constant $\boldsymbol{A} \in \mathbb{C}^{N\times M}$ dictionary matrix
$K \in \mathbb{N}$, with $K < N$, number of sources
$\text{SNR}_{\max} \in \mathbb{R}^+$ upper SNR limit in data
$\gamma_{\text{range}} \in [0,1]$, dynamic range for DOA grid pruning
$j_{\max} \in \mathbb{N}$ iteration count limit
$z \in \mathbb{N}$ with $z < j_{\max}$ convergence criterion
set $\boldsymbol{S_Y} = \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}/L$
set $\sigma^2_{\text{ceil}} = \text{tr}\,[\boldsymbol{S_Y}]$ upper limit on $\hat{\sigma}^2$
$\sigma^2_{\text{floor}} = \sigma^2_{\text{ceil}}/\text{SNR}_{\max}$ lower limit on $\hat{\sigma}^2$
initialize $\hat{\sigma}^2$, $\boldsymbol{\gamma}^{\text{new}}$, using Ref. [7, Eq.(24)], $j = 0$
repeat
$j = j + 1$ increment iteration counter
$\boldsymbol{\gamma}^{\text{old}} = \boldsymbol{\gamma}^{\text{new}}$
$\gamma_{\text{floor}} = \gamma_{\text{range}} \max(\boldsymbol{\gamma}^{\text{new}})$, source dynamic range
$\mathcal{P} = \{p \in \mathbb{N} \mid \gamma_p^{\text{new}} \geq \gamma_{\text{floor}}\}$ pruned DOA grid (16)
$\boldsymbol{\Gamma}_{\mathcal{P}} = \text{diag}(\boldsymbol{\gamma}_{\mathcal{P}}^{\text{new}})$ pruned source powers
$\boldsymbol{A}_{\mathcal{P}} = [\boldsymbol{a}_{p_1}, \ldots, \boldsymbol{a}_{p_P}]$ for all $p_i \in \mathcal{P}$, pruned dictionary
$\boldsymbol{\Sigma}_{\mathcal{P}} = \boldsymbol{A}_{\mathcal{P}}\boldsymbol{\Gamma}_{\mathcal{P}}\boldsymbol{A}_{\mathcal{P}}^{\mathsf{H}} + \hat{\sigma}^2\boldsymbol{I}_N$ (3)
$\boldsymbol{R_Y} = \frac{1}{Lb}\sum_{\ell=1}^{L} u(\boldsymbol{y}_\ell^{\mathsf{H}}\boldsymbol{\Sigma}_{\mathcal{P}}^{-1}\boldsymbol{y}_\ell;\cdot)\boldsymbol{y}_\ell\boldsymbol{y}_\ell^{\mathsf{H}}$ (12)
if using $u_{\text{Tyler}}(\cdot;\cdot)$ **then** normalize $\boldsymbol{R_Y} = \frac{\boldsymbol{R_Y}}{\text{tr}\,\boldsymbol{R_Y}}$
$\gamma_p^{\text{new}} = \gamma_p^{\text{old}} \left(\frac{\boldsymbol{a}_p^{\mathsf{H}}\boldsymbol{\Sigma}_{\mathcal{P}}^{-1}\boldsymbol{R_Y}\boldsymbol{\Sigma}_{\mathcal{P}}^{-1}\boldsymbol{a}_p}{\boldsymbol{a}_p^{\mathsf{H}}\boldsymbol{\Sigma}_{\mathcal{P}}^{-1}\boldsymbol{a}_p} \right)$ for all $p \in \mathcal{P}$ (13)
$\mathcal{M} = \{m \in \mathbb{N} \mid K \text{ largest peaks in } \boldsymbol{\gamma}^{\text{new}}\}$ active set
$\boldsymbol{A}_{\mathcal{M}} = [\boldsymbol{a}_{m_1}, \ldots, \boldsymbol{a}_{m_K}]$
$\hat{\sigma}^2_R = \frac{\text{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^{+})\boldsymbol{R_Y}\right]}{N-K}$ (14)
$\hat{\sigma}^2 = \max(\min(\hat{\sigma}^2_R, \sigma^2_{\text{ceil}}), \sigma^2_{\text{floor}})$ (15)
until ($\mathcal{M}$ has not changed during last z iterations) or $j > j_{\max}$
output $\mathcal{M}$ and j

Table 2. Robust Sparse Bayesian Learning for DOA [19].

Although the noise $\boldsymbol{n}$ is CES distributed, the snapshot data (signal + noise) is not. 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 2. The convergence criterion parameter $z = 10$ is chosen for all numerical simulations and the maximum number of iterations was set to $j_{\max} = 1200$, but this maximum was never reached.

Data $\boldsymbol{Y}$ are generated for the three source scenario ($K = 3$) with complex circularly symmetric zero-mean Gaussian amplitude from DOAs $\theta_{8701} = -3°$, $\theta_{9201} = 2°$, $\theta_{16501} = 75°$ according to the data models. The true active set $\mathcal{M}$ is $\{8701, 9201, 16501\}$ and source strengths are specified in the scenario as $\gamma_{8701} = \frac{1}{3}$, $\gamma_{9201} = \frac{1}{3}$, $\gamma_{16501} = \frac{1}{3}$ and $\gamma_m = 0$ for all $m \notin \mathcal{M}$, so that $\text{tr}\,[\boldsymbol{\Gamma}] = 1$.

scenario	DOAs	source variance
single source	$-10°$	$\gamma_{8001} = 1$
three sources	$-3°$, $2°$, $75°$	$\gamma_{8701} = \gamma_{9201} = \gamma_{16501} = \frac{1}{3}$

Table 3. Source scenarios, source variances normalized to $\text{tr}(\boldsymbol{\Gamma}) = 1$

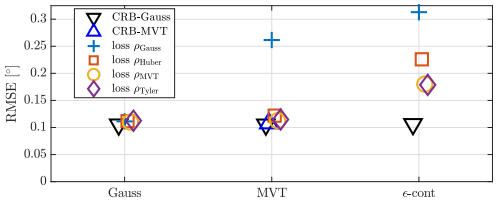

Fig. 1. RMSE for each DOA M-estimator at high ASNR = 30 dB for Gaussian, MVT, and ϵ-contaminated data.

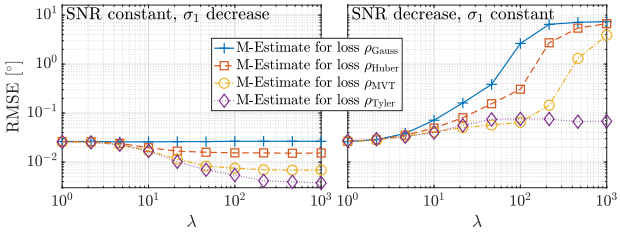

Fig. 2. For ϵ-contaminated data $\epsilon = 0.05$ with one source, RMSE versus outlier strength λ for each loss function for (left): ASNR= 25 dB and background noise σ_1 is decreasing, and (right): Background noise σ_1 is fixed and outlier noise $\lambda\sigma_1$ is increasing (at $\lambda = 1$: ASNR= 25 dB; and at $\lambda = 10^3$: ASNR $= 25 - 10\log(1 - \epsilon + \epsilon\lambda^2) = -22$ dB). RMSE evaluation based on $N_{\text{run}} = 250$ simulation runs.

The effect of the loss function on RMSE performance at high ASNR = 30 dB is illustrated in Fig. 1. This shows that for Gaussian data all choices of loss functions perform equally well at high ASNR. For MVT data in Fig. 1 (middle), we see that the robust loss functions (MVT, Huber, Tyler) work well, and approximately equally, whereas RMSE for Gauss loss is factor 2 worse. For ϵ-contaminated data in Fig. 1 (right) the Gauss loss performs a factor worse than the robust loss functions. Huber loss has slightly higher RMSE than MVT and Tyler loss.

For small outlier strength λ and for MVT data, the Gauss loss performs fine, but as the outlier noise increases the robust M-estimators outperform, see Fig. 2. As λ increases, the total noise changes, see (18). We here chose to keep the total noise constant in Fig. 2(left) by decreasing the background noise with increasing λ, or having the background noise constant in Fig. 2(right) whereby the total noise increases. For large outlier strength, Tyler loss performs best in Fig. 2(left) and does not breakdown in Fig. 2(right).

Due to the algorithmic speedup associated with the DOA grid pruning described in Sec. 3.4, it is feasible to run the algorithm in Table 2 with large dictionary size M with corre-

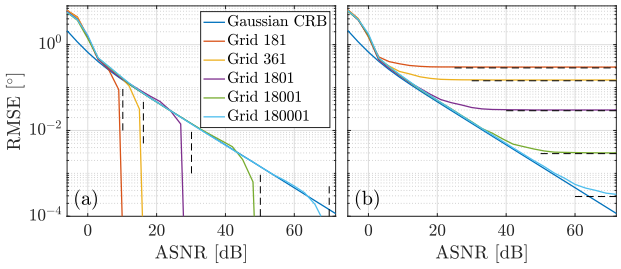

Fig. 3. Effect of dictionary size $M \in \{181, 361, 1801, 18001, 180001\}$ on RMSE vs. SNR for a single source a) fixed DOA $-10°$ on the grid and b) random uniformly distributed DOA $\sim -10° + U(-\delta/2, \delta/2)$. RMSE evaluation based on $N_{\text{run}} = 250$ simulation runs.

sponding angular grid resolution $\delta = 180°/(M-1)$.

The effect of grid resolution is illustrated in Fig. 3 for a single source impinging on a $N = 20$ element $\lambda/2$-spaced ULA. The Gaussian data model is used. Fig. 3 shows RMSE vs. ASNR for a dictionary size of $M \in \{181, 361, 1801, 18001, 180001\}$. In Fig. 3(a), the DOA is fixed at $-10°$, cf. single source scenario in Table 3, the DOA is on the angular grid which defines the dictionary matrix $\boldsymbol{A}$. In Fig. 3(b) the DOA is random, the source DOA is sampled from $-10° + U(-\delta/2, \delta/2)$ ($\delta = 180°/(M-1)$ is the angular grid resolution). The source DOA is not on the angular grid which defines the dictionary matrix $\boldsymbol{A}$.

For source DOA on the dictionary grid, Fig. 3(a), the RMSE performance resembles the behavior of an ML-estimator at low ASNR up to a certain threshold ASNR (dashed vertical lines) where the RMSE abruptly becomes zero. The threshold ASNR is deduced from the following argument: Let $\boldsymbol{a}_m$ be the true DOA dictionary vector and $\boldsymbol{a}_{m+1}$ be the dictionary vector for adjacent DOA on the angular grid. Comparing the corresponding Bartlett powers, we see that DOA errors become likely if the noise variance exceeds $2(|\boldsymbol{a}_m^{\mathsf{H}}\boldsymbol{a}_m| - |\boldsymbol{a}_m^{\mathsf{H}}\boldsymbol{a}_{m+1}|)/N = 2 - 2|\boldsymbol{a}_m^{\mathsf{H}}\boldsymbol{a}_{m+1}|/N$.

For source DOA off the dictionary grid, Fig. 3(b), the RMSE performance curve resembles the behavior of an ML-estimator at low ASNR up to a threshold ASNR. In the random DOA scenario, however, the RMSE flattens at increasing ASNR. Since the variance of the uniformly distributed source DOA is $\delta^2/12$, the limiting $\text{RMSE} = \delta/\sqrt{12}$ for $\text{ASNR} \to \infty$. The limiting RMSE (dashed horizontal lines) depends on the dictionary size M through the angular grid resolution δ. The asymptotic RMSE limits are shown as dashed horizontal lines in Fig. 3(b).

5. CONCLUSION

Robust Sparse Bayesian Learning (SBL) is derived based on loss functions and the assumption that the data are complex elliptically symmetric with finite covariances. The DOA M-estimator is available on GitHub [19]. The M-estimator's DOA RMSE is investigated numerically with synthetic array data for four loss functions: the ML-loss for the circular complex multivariate t-distribution (MVT), as well as Huber and Tyler loss. For Gauss loss, the method reduces to SBL. Synthetic array data are generated for Gaussian, MVT, and ϵ-contaminated models. The M-estimators perform well in simulations for MVT and ϵ-contaminated data and close to SBL for Gaussian data

6. REFERENCES

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

[2] 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.

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

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

[5] 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.

[6] 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.

[7] C.F. Mecklenbräuker, P. Gerstoft, and E. Ollila. DOA M-estimation using sparse Bayesian learning. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4933–4937, 2022.

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

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

[10] 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.

[11] U. K. Singh, R. Mitra, V. Bhatia, and A. K. Mishra. Kernel minimum error entropy based estimator for MIMO radar in non-Gaussian clutter. *IEEE Access*, 9:125320–125330, 2021.

[12] D. Luo, Z. Ye, B. Si, and J. Zhu. Deep MIMO radar target detector in Gaussian clutter. *IET Radar, Sonar & Navigation*, 2022.

[13] S. Feintuch, J. Tabrikian, I. Bilik, and H. Permuter. Neural network-based DOA estimation in the presence of non-Gaussian interference. *Trans. Aerosp. Electron. Syst.*, 2023. doi:10.1109/TAES.2023.3268256.

[14] J. T. Kent and D. E. Tyler. Redescending m-estimates of multivariate location and scatter. *Ann. Math. Statist.*, 19(4):2102–2119, 1991.

[15] E. Ollila, D. P. Palomar, and F. Pascal. Shrinking the eigenvalues of M-estimators of covariance matrix. *IEEE Trans. Signal Process.*, 69:256–269, 2020.

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

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

[18] P. J. Huber. Robust estimation of a location parameter. *Ann. Math. Statist.*, 35(1):73–101, 1964.

[19] Y. Park, E. Ollila, P. Gerstoft, and C.F. Mecklenbräuker. RobustSBL Repository. In `https://github.com/NoiseLabUCSD/RobustSBL`. GitHub, 2022.

[20] F. Meyer, P. Braca, P. Willett, and F. Hlawatsch. A scalable algorithm for tracking an unknown number of targets using multiple sensors. *IEEE Trans. Signal Process.*, 65(13):3478–3493, 2017.