Signal Processing 220 (2024) 109461

Contents lists available at ScienceDirect

Signal Processing

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

Robust and sparse M-estimation of DOA ☆

Christoph F. Mecklenbräuker [a,*], Peter Gerstoft [b], Esa Ollila [c], Yongsung Park [b]

[a] Institute of Telecommunications, TU Wien, 1040 Vienna, Austria
[b] University of California San Diego, La Jolla, CA 92093-0238, USA
[c] Dept. of Information and Communications Engineering, Aalto Univ. Aalto, Finland

ARTICLE INFO

Keywords:
DOA estimation
Robust statistics
Outliers
Sparsity
Complex elliptically symmetric
Bayesian learning

ABSTRACT

A robust and sparse Direction of Arrival (DOA) estimator is derived for array data that follows a Complex Elliptically Symmetric (CES) distribution with zero-mean and finite second-order moments. The derivation allows to choose the loss function and four loss functions are discussed in detail: the Gauss loss which is the Maximum-Likelihood (ML) loss for the circularly symmetric complex Gaussian distribution, the ML-loss for the complex multivariate t-distribution (MVT) with ν degrees of freedom, as well as Huber and Tyler loss functions. For Gauss loss, the method reduces to Sparse Bayesian Learning (SBL). The root mean square DOA error of the derived estimators is discussed for Gaussian, MVT, and ϵ-contaminated data. The robust SBL estimators perform well for all cases and nearly identical with classical SBL for Gaussian array data.

1. Introduction

Sparse processing for estimating DOAs is widely successful in finding a few active DOAs from sensor array data. An established method for estimating these DOAs is sparse Bayesian leaning which was derived assuming a Gaussian distribution for the sensor array data [1,2]. The Gaussian assumption might not always hold, thus we seek to estimate the DOAs assuming the array data follows a Complex Elliptically Symmetric (CES) distribution with zero-mean and finite second-order moments [3,4]. The CES assumption is widely used for non-Gaussian data, and well known to aid in the formulation of robust estimators.

Compressive beamformers for DOA estimation exhibit super-resolution properties even for a single measurement vector [5–7]. The cost functions employed by Robust and Sparse M-Estimation of DOA (as proposed in this work) are chosen to ensure that outliers in the data have limited influence on the DOA estimates.

SBL was derived under a complex multivariate Gaussian assumption on the data [8]. There is a rich literature on DOA estimation based on SBL for single and multiple measurements vectors [9–20]: SBL provides DOA estimates based on the sample covariance matrix (SCM) of the data sample. In Ref. [9] the SCM in the minimum variance adaptive beamformer was replaced with an estimate from a relevance vector machine. This improves its performance in case of source correlations or limited number of snapshots. An iterative algorithm using a Bayesian approach for an off-grid DOA model with joint sparsity among the snapshots was developed in Ref. [10].

Motivated by the use of elliptical distributions in generalized multivariate analysis for the study of distributions with heavy tails, we derive and investigate a robust statistical method that generalizes Sparse Bayesian Learning (SBL) for DOA from array data. We show that the DOA estimates are not unduly affected by outliers or other small departures from data distribution assumptions. The proposed method assumes CES data. In engineering applications, CES data models have been widely used when non-Gaussian models are needed [3,21–24].

Complex multinormality of the data may be a poorly fitting assumption. The noise may exhibit varying covariance over time, causing stationary data models to fit poorly over long observation times [25]. More advanced models are needed for applications in radar, sonar, mobile communication or imaging including the complex t-distribution, K-distribution, and generalized Gaussian [3]. The CES distribution [4] generalizes the complex multivariate normal distribution by allowing heavier or lighter tail probabilities in the data while maintaining the elliptical geometry of the equidensity contours. Elliptically symmetric distributions have been used in robust statistics to investigate how statistical procedures perform on data with outliers or heavy tails. Inference and performance bounds for parameter estimates from CES distributed data are found in [26–31].

SBL-type algorithms for DOA estimation in the presence of impulsive noise were proposed in Ref. [12]. This Bayes-optimal approach has high resolution and accuracy at the cost of high numerical complexity when the DOA grid is dense.

☆ This work was supported by Office of Naval Research, United States.
* Corresponding author.
E-mail address: cfm@tuwien.ac.at (C.F. Mecklenbräuker).

https://doi.org/10.1016/j.sigpro.2024.109461
Received 7 January 2024; Received in revised form 11 February 2024; Accepted 4 March 2024
Available online 7 March 2024

The SBL approach is flexible through the usage of various priors, although Gaussian are most common [2]. For Gaussian priors this has been approached based on minimization–majorization [7] and with expectation maximization (EM) [1,9,32–35]. We estimate the hyperparameters iteratively from the likelihood derivatives using stochastic maximum likelihood [36–38]. Recent investigations showed that SBL is lacking in statistical robustness [39–41]. A Bayes-optimal algorithm was proposed to estimate DOAs in the presence of impulsive noise from the perspective of SBL in [12].

1.1. Major contributions and outline

The major contributions of this paper are: (1) We formulate an SBL algorithm with a general loss function for DOA M-estimation which, given the number of sources, estimates the set of DOAs corresponding to non-zero source power. The data model parametrization is independent of the number of snapshots. (2) In addition to conventional Gaussian loss, we investigate the RMSE performance of DOA M-estimation for loss functions associated with data priors featuring potentially strong outliers (Huber, MVT and Tyler). This leads to a robust and sparse DOA M-estimator. RMSE performance is compared to the Cramér–Rao Bound (CRB) based on the Slepian–Bangs formula for Gaussian and CES data, which equals the semi-parametric CRB [28,29] for DOA estimation, cf. Appendix B. (3) The robust and sparse DOA M-estimator is shown to be insensitive to heavy tails, outliers, and unknown correlations among the sources.

The outline is as follows: Notation and the array data model are introduced in Section 2. Thereafter, we formulate the objective function and specific loss functions used for DOA M-estimator in Section 3 and describe the proposed algorithm. Simulation results for DOA estimation are discussed in Section 4 and report on convergence of the algorithm and associated run time in Section 5.

Notation

Matrices $\boldsymbol{A}, \boldsymbol{B}, \ldots$ and vectors $\boldsymbol{a}, \boldsymbol{b}, \ldots$ are complex-valued and denoted by boldface letters. The zero vector is $\boldsymbol{0}$, the $N \times N$ identity matrix is $\boldsymbol{I}_N$ and the diagonal matrix with $\boldsymbol{\gamma}$ on its main diagonal is $\mathrm{diag}(\boldsymbol{\gamma})$. The transpose, Hermitian transpose, trace, determinant, inverse, and Moore–Penrose pseudo inverse of $\boldsymbol{X}$ are denoted as $\boldsymbol{X}^{\mathsf{T}}, \boldsymbol{X}^{\mathsf{H}}, \mathrm{tr}\,[\boldsymbol{X}], \det \boldsymbol{X}, \boldsymbol{X}^{-1}, \boldsymbol{X}^{+}$ respectively. The sets $\mathbb{N}, \mathbb{R}$ and $\mathbb{R}^{+}$ denote the positive integers, reals and positive real numbers, respectively. The complex vector space of dimension N is written as $\mathbb{C}^N$ and the set of all complex-valued $N \times M$ matrices is $\mathbb{C}^{N\times M}$. The ℓ_p-norm is written as $\|\cdot\|_p$. The expectation of $\boldsymbol{X}$ is $\mathsf{E}(\boldsymbol{X})$ and if X and Y have the same distribution then we write $X \sim Y$.

2. Complex elliptically symmetric data model

Narrowband waves are observed on N sensors for L snapshots $\boldsymbol{y}_\ell$ and the data is $\boldsymbol{Y} = [\boldsymbol{y}_1 \ldots \boldsymbol{y}_L] \in \mathbb{C}^{N\times L}$. Here, the $\boldsymbol{y}_\ell \in \mathbb{C}^{N\times 1}$ are modeled as independent identically distributed (iid) CES,

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

where $\boldsymbol{0}$ is the mean, $\boldsymbol{\Sigma}$ is the scatter matrix and g is 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 [26,27].

The CES model (1) results from, e.g., a scale mixture of Gaussian distributions, which have a stochastic decomposition of the form

$$\boldsymbol{y}_\ell = \sqrt{\tau_\ell}\, \boldsymbol{v}_\ell,\ \text{with}\ \boldsymbol{v}_\ell = \boldsymbol{A}\boldsymbol{x}_\ell + \boldsymbol{n}_\ell, \quad (\ell = 1 \ldots L) \tag{3}$$

where $\tau_\ell > 0$ is a random variable independent of $\boldsymbol{v}_\ell$. 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 $\boldsymbol{n}_\ell \sim \mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$. Eq. (3) models a sub class of the general model (2) and serves as a computational model for synthetic data generation in Sections 4 and 5.

The random variables τ_ℓ enable to accommodate various array data distributions in the model. The CES data model (1) includes complex Gaussian array data as a special case and generalizes this by allowing higher or lower tail probabilities than the complex Gaussian while keeping the elliptical contours of equal density [3].

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. [41,42]. The scatter matrix is then modeled as

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{H}} + \sigma^2 \boldsymbol{I}_N, \tag{4}$$

$$\boldsymbol{\Gamma} = \mathrm{cov}(\boldsymbol{x}_\ell) = \mathrm{diag}(\boldsymbol{\gamma}) \tag{5}$$

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

The unknown zero-mean complex source amplitudes are the elements of $\boldsymbol{X} = [\boldsymbol{x}_1 \ldots \boldsymbol{x}_L] \in \mathbb{C}^{M\times L}$ where M is the considered number of hypothetical DOAs on the given grid $\{\theta_1, \ldots, \theta_M\}$. The source amplitudes are independent across sources and snapshots, i.e., x_{ml} and $x_{m'l'}$ are independent for $(m, l) \neq (m', l')$. If K sources are present in the ℓth array data snapshot, the ℓth column of $\boldsymbol{X}$ is K-sparse and we assume that the sparsity pattern is the same for all snapshots. The sparsity pattern is modeled by the active set

$$\mathcal{M} = \big\{m \in \{1, \ldots, M\} \mid \gamma_m \neq 0\big\} = \{m_1, \ldots, m_K\}, \tag{6}$$

and $x_{m\ell} = 0$ for all $m \notin \mathcal{M}$. The noise $\boldsymbol{N} = [\boldsymbol{n}_1 \ldots \boldsymbol{n}_L] \in \mathbb{C}^{N\times L}$ is assumed iid across sensors and snapshots, zero-mean, with finite variance σ^2 for all n, ℓ.

It follows from (3) that $\boldsymbol{y}_\ell | \tau_\ell \sim \mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \tau_\ell \boldsymbol{\Sigma})$. The so-called *density generator* $g(\cdot)$ [3],[4, Eq.(4.15)] is evaluated by

$$g(t) = \pi^{-N} \int_0^\infty \tau^{-N} e^{-t/\tau} p_\tau(\tau)\, \mathrm{d}\tau. \tag{7}$$

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

2.1. Dictionary for plane wave arrivals

The M columns of the dictionary $\boldsymbol{A} = [\boldsymbol{a}_1 \ldots \boldsymbol{a}_M]$ are the replica vectors for all hypothetical DOAs. For a uniform linear array (ULA), the dictionary matrix elements are $A_{nm} = \mathrm{e}^{-\mathrm{j}(n-1)\frac{2\pi d}{\lambda}\sin\theta_m}$ where λ is the wavelength and d is the element spacing.

The dictionary's DOA grid is defined as $\theta_m = -90° + (m-1)\delta$, $\forall m = 1, \ldots, M$ where δ is the dictionary's angular grid resolution, $\delta = 180°/(M-1)$.

The K "active" replica vectors are aggregated in

$$\boldsymbol{A}_{\mathcal{M}} = [\boldsymbol{a}_{m_1} \ldots \boldsymbol{a}_{m_K}] \in \mathbb{C}^{N\times K}, \tag{8}$$

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

2.2. Traditional signal plus noise model

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

$$\boldsymbol{y}_\ell = \boldsymbol{A}\boldsymbol{x}_\ell + \boldsymbol{n}_\ell\ . \tag{9}$$

This model results in Gaussian data, $\boldsymbol{y}_\ell \sim \mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \boldsymbol{\Sigma})$. Model (9) was assumed in [11] 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) [43,44] 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 [4,40]. The complex MVT distribution is a special case of the CES distribution, cf. Table 1.

3. M-estimation based on CES distribution

3.1. Covariance matrix objective function

We follow a general approach based on loss functions and assume that the array data $\boldsymbol{Y}$ has a CES distribution with zero mean $\mathbf{0}$ and positive definite Hermitian $N \times N$ covariance matrix parameter $\boldsymbol{\Sigma}$ [3,45]. Thus

$$p(\boldsymbol{Y}|\mathbf{0}, \boldsymbol{\Sigma}) = C^L \prod_{\ell=1}^{L} \det(\boldsymbol{\Sigma}^{-1}) g(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell). \tag{10}$$

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 [4, (4.20)],

$$\mathcal{L}(\boldsymbol{\Sigma}) = \frac{1}{Lb} \sum_{\ell=1}^{L} \rho(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell) - \log\det(\boldsymbol{\Sigma}^{-1}), \tag{11}$$

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. [4, Sec. 4.3]. A specific choice of loss function ρ renders (11) equal to the negative log-likelihood of $\boldsymbol{\Sigma}$ when the array data are CES distributed with density generator $g(t) = \mathrm{e}^{-\rho(t)}$ [46]. If the loss function is chosen, e.g., as $\rho(t) = t$ then (11) becomes the negative log-likelihood function for $\boldsymbol{\Sigma}$ in the Gaussian model (9).

The term b is a fitting coefficient, called consistency factor, which makes the minimizer of (11) consistent to $\boldsymbol{\Sigma}$ when the array data are Gaussian [4, Sec. 4.3]. Thus, $\boldsymbol{y} \sim \mathbb{C}\mathcal{N}_N(\mathbf{0}, \boldsymbol{I})$,

$$b = \mathsf{E}[\psi(\|\boldsymbol{y}\|_2^2)]/N, \tag{12}$$

$$= \frac{1}{N} \int_0^\infty \psi(t/2) f_{\chi^2_{2N}}(t) \mathrm{d}t \tag{13}$$

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. To arrive from (12) to (13) we used $\|\boldsymbol{y}\|^2 \sim (1/2)\chi^2_{2N}$.

Minimizing (11) with b according to (13) results in a consistent M-estimator of the covariance matrix $\boldsymbol{\Sigma}$ when the objective function is derived under a given non-Gaussian array data assumption (as in Section 3.2) but is in fact Gaussian ($\boldsymbol{y}_\ell \sim \mathbb{C}\mathcal{N}_N(\mathbf{0}, \boldsymbol{\Sigma})$). Further details are given in Appendix A.

3.2. Loss functions

We discuss four different choices of loss function $\rho(\cdot)$. These loss functions are summarized in Table 1. For each loss function, we also summarize the consistency factor b and the weight function $u(t) = \mathrm{d}\rho(t)/\mathrm{d}t$ associated with the loss function ρ.

Gauss loss corresponds to loss function of (circular complex) Gaussian distribution:

$$\rho_{\text{Gauss}}(t) = t \tag{14}$$

in which case the consistency factor $b = 1$ and the objective in (11) becomes the Gaussian (negative) log-likelihood function $\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_{\boldsymbol{Y}}\} - \log\det(\boldsymbol{\Sigma}^{-1})$ where

$$\boldsymbol{S}_{\boldsymbol{Y}} = \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}/L, \tag{15}$$

is the SCM and the minimizer of which is $\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}_{\boldsymbol{Y}}$. In this case b in (13) becomes $b = 1$ as expected, since $\boldsymbol{S}_{\boldsymbol{Y}}$ is consistent to $\boldsymbol{\Sigma}$ without any scaling correction. For Gauss loss $u_{\text{Gauss}}(t) = 1$.

Huber loss is given by [4, Eq. (4.29)]

$$\rho_{\text{Huber}}(t; c) = \begin{cases} t & \text{for } t \leqslant c^2, \\ c^2\left(\log(t/c^2) + 1\right) & \text{for } t > c^2. \end{cases} \tag{16}$$

The threshold c is a tuning parameter that affects the robustness and efficiency of the estimator. Huber loss specializes the objective function (11) to the negative log-likelihood of $\boldsymbol{\Sigma}$ when the array data are heavy-tailed CES distributed with a density generator of the form $\mathrm{e}^{-\rho_{\text{Huber}}(t;c)}$. The squared threshold c^2 in (16) is mapped to the qth quantile of $(1/2)\chi^2_{2N}$-distribution and we regard $q \in (0, 1)$ as a loss parameter which is chosen by design, see Table 1. It is easy to verify that b in (13) for Huber loss function is [4, Sec. 4.4.2],

$$\begin{aligned} b_{\text{Huber}} &= F_{\chi^2_{2(N+1)}}(2c^2) + c^2(1 - F_{\chi^2_{2N}}(2c^2))/N, \\ &= F_{\chi^2_{2(N+1)}}(2c^2) + c^2(1 - q)/N, \end{aligned} \tag{17}$$

where $F_{\chi^2_{2N}}(x)$ denotes the cumulative distribution of the χ^2_{2N} distribution.

For Huber loss (16) the weight function becomes

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

Thus, an observation $\boldsymbol{y}_\ell$ with squared Mahalanobis distance (MD) $\boldsymbol{y}_\ell^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell$ smaller than c^2 receives constant weight, while observations with a larger MD are heavily down-weighted.

MVT loss corresponds to the ML-loss for (circular complex) multivariate t (MVT) distribution with ν_{loss} degrees of freedom, $\boldsymbol{y}_i \sim \mathbb{C}t_{N,\nu}(\mathbf{0}, \boldsymbol{\Sigma})$ [4, Eq.(4.28)],

$$\rho_{\text{MVT}}(t; \nu_{\text{loss}}) = \frac{\nu_{\text{loss}} + 2N}{2} \log(\nu_{\text{loss}} + 2t). \tag{19}$$

The ν_{loss} parameter in (19) is viewed as a loss parameter which is chosen by design, see Table 1. The consistency factor b_{MVT} for $\rho_{\text{MVT}}(t; \nu_{\text{loss}})$ is computed numerically

For MVT-loss (19) the corresponding weight is

$$u_{\text{MVT}}(t; \nu_{\text{loss}}) = \frac{\nu_{\text{loss}} + 2N}{\nu_{\text{loss}} + 2t}, \tag{20}$$

Tyler loss is given by [47], [4, Sec. 4.4.3, Eq. (4.30)]

$$\rho_{\text{Tyler}}(t) = N \log(t). \tag{21}$$

which is the limiting case of $\rho_{\text{MVT}}(t; \nu_{\text{loss}})$ for $\nu_{\text{loss}} \to 0$ and of $\rho_{\text{Huber}}(t; c)$ for $c \to 0$. This is the ML-loss of the Angular Central Gaussian (ACG) distribution [4, Sec. 4.2.3] which is not a CES distribution. For Tyler loss (21) the weight function becomes

$$u_{\text{Tyler}}(t) = N/t. \tag{22}$$

In this case, Tyler's M-estimator estimates the shape of the covariance matrix $\boldsymbol{\Sigma}$ only. Namely, for Tyler loss and $b = 1$, if $\hat{\boldsymbol{\Sigma}}$ is a minimizer of (11) then so is $b\hat{\boldsymbol{\Sigma}}$ for any $b > 0$. Thus, the solution is unique only up to a scale.

For Tyler loss $\psi(t) \equiv N\ \forall t$, indicating that the consistency factor for Tyler loss cannot be found based on (12). In this case, $\boldsymbol{R}_{\boldsymbol{Y}}$ is defined as weighted SCM that uses normalized Tyler weights (A.6), see Appendix A.

Table 1
Density generators used for CES data generation, loss and weight functions used in DOA M-estimation.

Loss name	Data density generator $g(t)$	Loss $\rho(t)$	Weight $u(t)$	$\psi(t)$	Loss param.
	Eq. (2)	$-\log g(t)$	$\rho'(t)$	$tu(t)$	
Gauss	e^{-t}	t	1	t	n/a
MVT	$(1 + t/\nu)^{-(\nu+2N)/2}$	$\frac{\nu+2N}{2}\log(\nu + 2t)$	$\frac{\nu+2N}{\nu+2t}$	$\frac{\nu+2N}{2+\nu/t}$	$\nu = 2.1$
Huber	$\exp(-\rho_{\text{Huber}}(t; c))$	[4, Eq. (4.29)]	$= \begin{cases} 1 \text{ if } t < c^2 \\ c^2/t \text{ else} \end{cases}$	$t\,u_{\text{Huber}}(t)$	$q = 0.9$
Tyler	t^{-N}	$N \log t$	N/t	N	n/a

3.3. Source power estimation

Similarly to Ref. [11, Sec. III.D], we regard (11) as a function of $\boldsymbol{\gamma}$ and σ^2 and compute the first order derivative

$$\frac{\partial \mathcal{L}}{\partial \gamma_m} = \boldsymbol{a}_m^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m - \frac{1}{Lb} \sum_{\ell=1}^{L} \|\boldsymbol{a}_m^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell\|_2^2 u(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell). \tag{23}$$

Eq. (23) is identical to Ref. [11, Eq.(21)] except for the weight function $u(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell)$. For the Gaussian array data model, the weight function is the constant function $u_{\text{Gauss}}(t) \equiv 1$.

Setting (23) to zero gives

$$\boldsymbol{a}_m^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m = \boldsymbol{a}_m^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{R}_{\boldsymbol{Y}} \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m, \ \forall m \in \{1, \dots, M\}, \tag{24}$$

where $\boldsymbol{R}_{\boldsymbol{Y}}$ is the weighted SCM

$$\boldsymbol{R}_{\boldsymbol{Y}} = \frac{1}{Lb} \sum_{\ell=1}^{L} u(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell) \boldsymbol{y}_\ell \boldsymbol{y}_\ell^{\mathsf{H}} = \frac{1}{L} \boldsymbol{Y} \boldsymbol{D} \boldsymbol{Y}^{\mathsf{H}} \tag{25}$$

with $\boldsymbol{D} = \operatorname{diag}(u_1, \dots, u_L)/b$ and $u_\ell = u(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell)$. Note that $\boldsymbol{R}_{\boldsymbol{Y}}$ can be understood as an adaptively weighted SCM [4, Sec. 4.3]. $\boldsymbol{R}_{\boldsymbol{Y}}$ is Fisher consistent for the covariance matrix when $\boldsymbol{y}_\ell$ are Gaussian distributed, i.e., $\mathsf{E}[\boldsymbol{R}_{\boldsymbol{Y}}] = \boldsymbol{\Sigma}$ thanks to the consistency factor b [4, Sec. 4.4.1]. Although (24) is rather concise, it hides the parameters $\boldsymbol{\gamma}, \sigma^2$ in $\boldsymbol{\Sigma}$ and $\boldsymbol{R}_{\boldsymbol{Y}}$ and a closed-form solution for $\boldsymbol{\gamma}, \sigma^2$ is not known to the authors. Therefore, we formulate a fixed-point equation for each γ_m given σ^2 and γ_p, $p \neq m \in \{1, \dots, M\}$ for solving (24) by iterations. Let γ_m^{old} be a previous approximation for γ_m then (similarly to EM-based approaches [1, Eqs. (18)–(19)] and [48]) the gradient-based update is computed with stepsize μ, cf. [49, Ch. 8]

$$\gamma_m^{\text{new}} = (1-\mu)\gamma_m^{\text{old}} + \mu \gamma_m^{\text{old}} G_m(\boldsymbol{\gamma}^{\text{old}}), \quad \forall m, \tag{26}$$

where $0 < \mu \leq 1$ and

$$G_m(\boldsymbol{\gamma}) = \frac{\boldsymbol{a}_m^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{R}_{\boldsymbol{Y}} \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m}{\boldsymbol{a}_m^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_m}, \quad \forall m. \tag{27}$$

The definition of $G_m(\boldsymbol{\gamma})$ in (27) ensures that (24) is fulfilled at the fixed-point of (26). In (27), both $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{R}_{\boldsymbol{Y}}$ depend on $\boldsymbol{\gamma}$. The simulations results in Section 4 are obtained for stepsize $\mu = 1$. Convergence, iteration count and run-time are discussed in Section 5.

The active set $\mathcal{M}$ is selected as either the K largest entries of $\boldsymbol{\gamma}$ or the entries with γ_m exceeding a threshold.

3.4. Noise variance estimation

The original SBL algorithm exploits Jaffer's necessary condition [37, Eq. (6)] which leads to the noise subspace based estimate [34, Eq. (15)], [11, Sec. III.E],

$$\hat{\sigma}_S^2 = \operatorname{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}} \boldsymbol{A}_{\mathcal{M}}^{+}) \boldsymbol{S}_{\boldsymbol{Y}}\right] \Big/ (N-K). \tag{28}$$

This noise variance estimate works well with DOA estimation [11,48, 50] without outliers in the array data. The derivation of the robust noise variance estimate starts from the following robust formulation of Jaffer's condition, namely

$$\boldsymbol{A}_{\mathcal{M}}^{\mathsf{H}} (\boldsymbol{R}_{\boldsymbol{Y}} - \boldsymbol{\Sigma}) \boldsymbol{A}_{\mathcal{M}} = \boldsymbol{0}. \tag{29}$$

This results in the robust estimate

$$\hat{\sigma}_R^2 = \operatorname{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}} \boldsymbol{A}_{\mathcal{M}}^{+}) \boldsymbol{R}_{\boldsymbol{Y}}\right] \Big/ (N-K), \tag{30}$$

for the full derivation see [40]. For Gauss loss function, $\boldsymbol{R}_{\boldsymbol{Y}} = \boldsymbol{S}_{\boldsymbol{Y}}$, and the expressions (28), (30) are identical.

To stabilize the noise variance M-estimate (30) for non-Gauss loss, we define lower and upper bounds for $\hat{\sigma}^2$ and enforce $\sigma_{\text{floor}}^2 \leq \hat{\sigma}^2 \leq \sigma_{\text{ceil}}^2$ by

$$\hat{\sigma}^2 = \max(\min(\hat{\sigma}_R^2, \sigma_{\text{ceil}}^2), \sigma_{\text{floor}}^2) \tag{31}$$

Table 2
Robust and Sparse M-Estimation of DOA for single-frequency array data. Matlab function `SBL_v5p12.m` [51] implements Robust and Sparse M-Estimation of DOA from multi-frequency array data.

input $\boldsymbol{Y} \in \mathbb{C}^{N \times L}$ array data to be analyzed
select the weight function $u(\cdot)$ and loss parameter
constant $\boldsymbol{A} \in \mathbb{C}^{N \times M}$ dictionary matrix
$K \in \mathbb{N}$, with $K < N$, number of sources
$\delta \in \mathbb{R}^+$ small positive constant
$\text{SNR}_{\max} \in \mathbb{R}^+$ upper SNR limit in data
$\gamma_{\text{range}} \in [0, 1]$, dynamic range for grid pruning
$\mu \in [0, 1]$, iteration stepsize
$j_{\max} \in \mathbb{N}$ iteration count limit
$z \in \mathbb{N}$ with $z < j_{\max}$ convergence criterion
set $\boldsymbol{S}_{\boldsymbol{Y}} = \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}/L$
set $\sigma_{\text{ceil}}^2 = \operatorname{tr}\left[\boldsymbol{S}_{\boldsymbol{Y}}\right]$ upper limit on $\hat{\sigma}^2$
$\sigma_{\text{floor}}^2 = \sigma_{\text{ceil}}^2/\text{SNR}_{\max}$ lower limit on $\hat{\sigma}^2$
initialize $\hat{\sigma}^2$, $\boldsymbol{\gamma}^{\text{new}}$ using (33)–(36), $j = 0$
repeat
  $j = j + 1$ iteration counter
  $\boldsymbol{\gamma}^{\text{old}} = \boldsymbol{\gamma}^{\text{new}}$
  $\gamma_{\text{floor}} = \gamma_{\text{range}} \max(\boldsymbol{\gamma}^{\text{new}})$, source dynamic range
  $\mathcal{P} = \{p \in \mathbb{N} \mid \boldsymbol{\gamma}_p^{\text{new}} \geq \gamma_{\text{floor}}\}$ pruned grid (32)
  $\boldsymbol{\Gamma}_{\mathcal{P}} = \operatorname{diag}(\boldsymbol{\gamma}_{\mathcal{P}}^{\text{new}})$ pruned source powers
  $\boldsymbol{A}_{\mathcal{P}} = [\boldsymbol{a}_{p_1}, \dots, \boldsymbol{a}_{p_P}]$ for all $p_i \in \mathcal{P}$, pruned dictionary
  $\boldsymbol{\Sigma}_{\mathcal{P}} = \boldsymbol{A}_{\mathcal{P}} \boldsymbol{\Gamma}_{\mathcal{P}} \boldsymbol{A}_{\mathcal{P}}^{\mathsf{H}} + \hat{\sigma}^2 \boldsymbol{I}_N$ (4)
  if using $u_{\text{Tyler}}(\cdot)$ **then** $b = \frac{1}{L} \sum_{\ell=1}^{L} u_{\text{Tyler}}(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}_\ell)$ (A.6)
  $\boldsymbol{R}_{\boldsymbol{Y}} = \frac{1}{Lb} \sum_{\ell=1}^{L} u(\boldsymbol{y}_\ell^{\mathsf{H}} \boldsymbol{\Sigma}_{\mathcal{P}}^{-1} \boldsymbol{y}_\ell) \boldsymbol{y}_\ell \boldsymbol{y}_\ell^{\mathsf{H}}$ (25)
  $\gamma_p^{\text{new}} = (1-\mu)\gamma_p^{\text{old}} + \mu \gamma_p^{\text{old}} G_p(\boldsymbol{\gamma}^{\text{old}})$, for all $p \in \mathcal{P}$ (26)
  $\mathcal{M} = \{m \in \mathbb{N} \mid K$ largest peaks in $\boldsymbol{\gamma}^{\text{new}}\}$ active set
  $\boldsymbol{A}_{\mathcal{M}} = [\boldsymbol{a}_{m_1}, \dots, \boldsymbol{a}_{m_K}]$
  $\hat{\sigma}_R^2 = \operatorname{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}} \boldsymbol{A}_{\mathcal{M}}^{+}) \boldsymbol{R}_{\boldsymbol{Y}}\right] \Big/ (N-K)$ (30)
  $\hat{\sigma}^2 = \max(\min(\hat{\sigma}_R^2, \sigma_{\text{ceil}}^2), \sigma_{\text{floor}}^2)$ (31)
until ($\mathcal{M}$ has not changed in last z iterations) or $j > j_{\max}$
output $\mathcal{M}$ and j

The original SBL algorithm [11] does not use/need this stabilization because for Gauss loss, the weighted SCM (25) equals (15) which does not depend on prior knowledge of $\boldsymbol{\Sigma}$. We have chosen $\sigma_{\text{floor}}^2 = 10^{-6} \operatorname{tr}[\boldsymbol{S}_{\boldsymbol{Y}}]/N$ and $\sigma_{\text{ceil}}^2 = \operatorname{tr}[\boldsymbol{S}_{\boldsymbol{Y}}]/N$ for the simulations.

As discussed in Appendix A, 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.

3.5. Algorithm

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

DOA grid pruning. To reduce numerical complexity in the iterations, we introduce the pruned DOA grid $\mathcal{P}$ by not wasting computational resources on those DOAs which are associated with source power estimates below a chosen threshold value γ_{floor}, i.e., we introduce a thresholding operation on the $\boldsymbol{\gamma}^{\text{new}}$ vector. The pruned DOA grid is formally defined as an index set,

$$\mathcal{P} = \{p \in \{1, \dots, M\} \mid \boldsymbol{\gamma}_p^{\text{new}} \geq \gamma_{\text{floor}}\} = \{p_1, \dots, p_P\}. \tag{32}$$

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

Initialization. In our algorithm we need to give initial values of source signal powers $\boldsymbol{\gamma}$ and the noise variance σ^2. The initial values are computed via following steps:

(a) Compute $\boldsymbol{S}_{\boldsymbol{Y}}$ and CBF output powers

$$\gamma_m^{\text{init}} = \frac{\boldsymbol{a}_m^{\mathsf{H}} \boldsymbol{S}_{\boldsymbol{Y}} \boldsymbol{a}_m}{\|\boldsymbol{a}_m\|_2^4}, \quad \forall m = 1, \dots, M. \tag{33}$$

(b) Compute the initial active set by identifying K largest peaks in the CBF output powers,

$$\mathcal{M} = \{m \in \mathbb{N} \mid K \text{ largest peaks in } \boldsymbol{\gamma}^{\text{init}}\} \tag{34}$$

(c) Compute the initial noise variance

$$\hat{\sigma}^2 = \hat{\sigma}_S^2 = \frac{\operatorname{tr}\left[(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^{+})\boldsymbol{S}_{\boldsymbol{Y}}\right]}{N-K} \tag{35}$$

(d) Compute initial estimates of source powers:

$$\gamma_m^{\text{new}} = \max(\delta, (\gamma_m^{\text{init}} - \hat{\sigma}^2)), \quad \forall m = 1, \dots, M. \tag{36}$$

where $\delta > 0$ is a small number, guaranteeing that all initial γ_m^{new} are positive.

3.5.1. Convergence criterion

The DOA estimates returned by the algorithm in Table 2 are obtained from the active set $\mathcal{M}$. Therefore, the active set is monitored for changes 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) terminates. 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

The proposed DOA M-estimation algorithm using SBL is displayed in Table 2. 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 three scenarios with $K = 1, \dots, 3$ incoming plane waves and corresponding DOAs as listed in Table 3. The source amplitudes $\boldsymbol{x}_\ell$ are complex circularly symmetric zero-mean Gaussian. The wavefield is modeled according to (1) and it is observed by a uniform linear array with $N = 20$ elements at half-wavelength spacing. The dictionary $\boldsymbol{A}$ consists of $M = 18001$ replica vectors for the high-resolution DOA grid $\theta_m = -90° + (m-1)\delta$, $\forall m = 1, \dots, M$ where δ is the dictionary's angular grid resolution, $\delta = 180°/(M-1) = 0.01°$.

The convergence criterion parameter $z = 10$ is chosen for all simulations and the maximum number of iterations was set to $j_{\max} = 1200$, but this was never reached. The RMSE of the DOA estimates over $N_{\text{run}} = 250$ simulation runs with random array data realizations is used for evaluating the performance of the algorithm,

$$\text{RMSE} = \sqrt{\sum_{r=1}^{N_{\text{run}}} \sum_{k=1}^{K} \frac{[\min(|\hat{\theta}_k^r - \theta_k^r|, e_{\max})]^2}{K\, N_{\text{run}}}}, \tag{37}$$

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. [52]. We use $e_{\max} = 10°$ in (37). Thus, maximum RMSE is 10°.

Data generation. The array data are generated according to three models which are listed in Table 4 and explained below:

Gaussian array data. In this model $\boldsymbol{y}_\ell = \boldsymbol{v}_\ell \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ from (4).

MVT array data. We draw $\boldsymbol{v}_\ell \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $s_\ell \sim \chi^2_{\nu_{\text{data}}}$ independently, with $\boldsymbol{\Sigma}$ from (4). We set $\tau_\ell = \nu_{\text{data}}/s_\ell$ and the array data is modeled by the scale mixture $\boldsymbol{y}_\ell = \sqrt{\tau_\ell}\boldsymbol{v}_\ell \sim \mathbb{C}t_{\nu_{\text{data}}}$-distributed, cf. [39] and [4, Sec. 4.2.2].

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

Scenario	DOAs	Source variance
Single source	$-10°$	$\gamma_{8001} = 1$
Two sources	$-10°$, $10°$	$\gamma_{8001} = \gamma_{10001} = \frac{1}{2}$
Three sources	$-3°$, $2°$, $75°$	$\gamma_{8701} = \gamma_{9201} = \gamma_{16501} = \frac{1}{3}$

Table 4
Array data models.

Array data model	Eq.	Parameters	ASNR
Gaussian	(9)	$\boldsymbol{\Sigma}$	N/σ^2
MVT	(1)	$\boldsymbol{\Sigma}$, $\nu_{\text{data}} = 2.1$	N/σ^2
ϵ-contaminated	(9)	$\boldsymbol{\Sigma}$, $\epsilon = 0.05$, $\lambda = 10$	N/σ^2

ϵ-Contaminated array data. This heavy-tailed array data model is not covered by (1) with the assumptions in Section 2. Instead, the noise $\boldsymbol{n}$ is drawn with probability $(1-\epsilon)$ from $\mathbb{C}\mathcal{N}(\boldsymbol{0}, \sigma_1^2\boldsymbol{I})$ and with probability ϵ from a $\mathbb{C}\mathcal{N}(\boldsymbol{0}, \lambda^2\sigma_1^2\boldsymbol{I})$, where λ is the outlier strength. Thus, $\boldsymbol{y}_\ell$ is drawn from $\mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{H}} + \sigma_1^2\boldsymbol{I}_N)$, using (4) with probability $(1-\epsilon)$ and with outlier probability ϵ from $\mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{H}} + (\lambda\sigma_1)^2\boldsymbol{I}_N)$. The resulting noise covariance matrix is $\sigma^2\boldsymbol{I}_N$ similar to the other models, but with

$$\sigma^2 = (1 - \epsilon + \epsilon\lambda^2)\sigma_1^2. \tag{38}$$

The limiting distribution of ϵ-contaminated noise for $\epsilon \to 0$ and any constant $\lambda > 0$ is Gaussian.

Additionally, the CRB for DOA estimation from Gaussian and MVT array data are shown. The CRB is a lower bound and provided for reference in the plots. The CRB is not tight for low SNR and a tighter lower bound for DOA for both high and low SNR is the Ziv–Zakai bound [53] which could potentially be formulated for the source scenarios and array data models in Tables 3 and 4. The corresponding expressions are given in Appendix B for completeness. The Gaussian CRB shown in Figs. 1(a,d,g) is evaluated according to [54, Eq. (8.106)]. The bound for MVT array data, $C_{\text{CR,MVT}}(\boldsymbol{\theta})$, is just slightly higher than the bound for Gaussian array data, $C_{\text{CR,Gauss}}(\boldsymbol{\theta})$. For the three source scenario, $N = 20$, $L = 25$, Gaussian and MVT array data model, the gap between the bounds $C_{\text{CR,Gauss}}(\boldsymbol{\theta})$ and $C_{\text{CR,MVT}}(\boldsymbol{\theta})$ is smaller than 3% in terms of RMSE in the shown ASNR range. The gap is not observable in the RMSE plots in Fig. 1 for the chosen ASNR and RMSE ranges. The MVT CRB is shown in Figs. 1(b,e,h). We have not evaluated the CRB for ϵ-contaminated array data. In the performance plots for ϵ-contaminated array data, we show $C_{\text{CR,Gauss}}(\boldsymbol{\theta})$ for ASNR $= N/\sigma^2$ using (38), labeled as "Gaussian CRB (shifted)" in Figs. 1(c,f,i). We expect that the true CRB for ϵ-contaminated array data is higher than this approximation.

Single source scenario. A single plane wave ($K = 1$) with complex circularly symmetric zero-mean Gaussian amplitude is arriving from DOA $\theta_{8001} = -10°$. Here, ASNR $= N/\sigma^2$, cf. [54, Eq. (8.112)]. Fig. 1 shows results for RMSE of DOA estimates in scenarios with $L = 25$ snapshots and $N = 20$ sensors. RMSE is averaged over 250 iid realizations of DOA estimates from array data $\boldsymbol{Y}$. There are more snapshots L than sensors N, ensuring full rank $\boldsymbol{R}_{\boldsymbol{Y}}$ almost surely.

Simulations for Gaussian noise are shown in Fig. 1(a). For this scenario, the conventional beamformer (not shown) is the ML DOA estimator and approaches the CRB for ASNR greater than 3 dB. All shown M-estimators for DOA perform almost identically in terms of RMSE and just slightly worse than the CBF.

Fig. 1(b) shows simulations for heavy-tailed MVT distributed array data with degrees of freedom parameter $\nu_{\text{data}} = 2.1$. We observe that the M-estimator for MVT-loss ρ_{MVT} for loss parameter $\nu_{\text{loss}} = 2.1$ performs the best, closely followed by the M-estimators for Tyler loss and Huber-loss ρ_{Huber} for loss parameter $q = 0.9$. Here, the loss parameter ν_{loss} used by M-estimator is identical to the MVT array data parameter ν_{data} and, thus, is expected to work well. In Fig. 1(b), the M-estimator for MVT-loss ρ_{MVT} closely follows the corresponding MVT CRB (B.1) for ASNR >

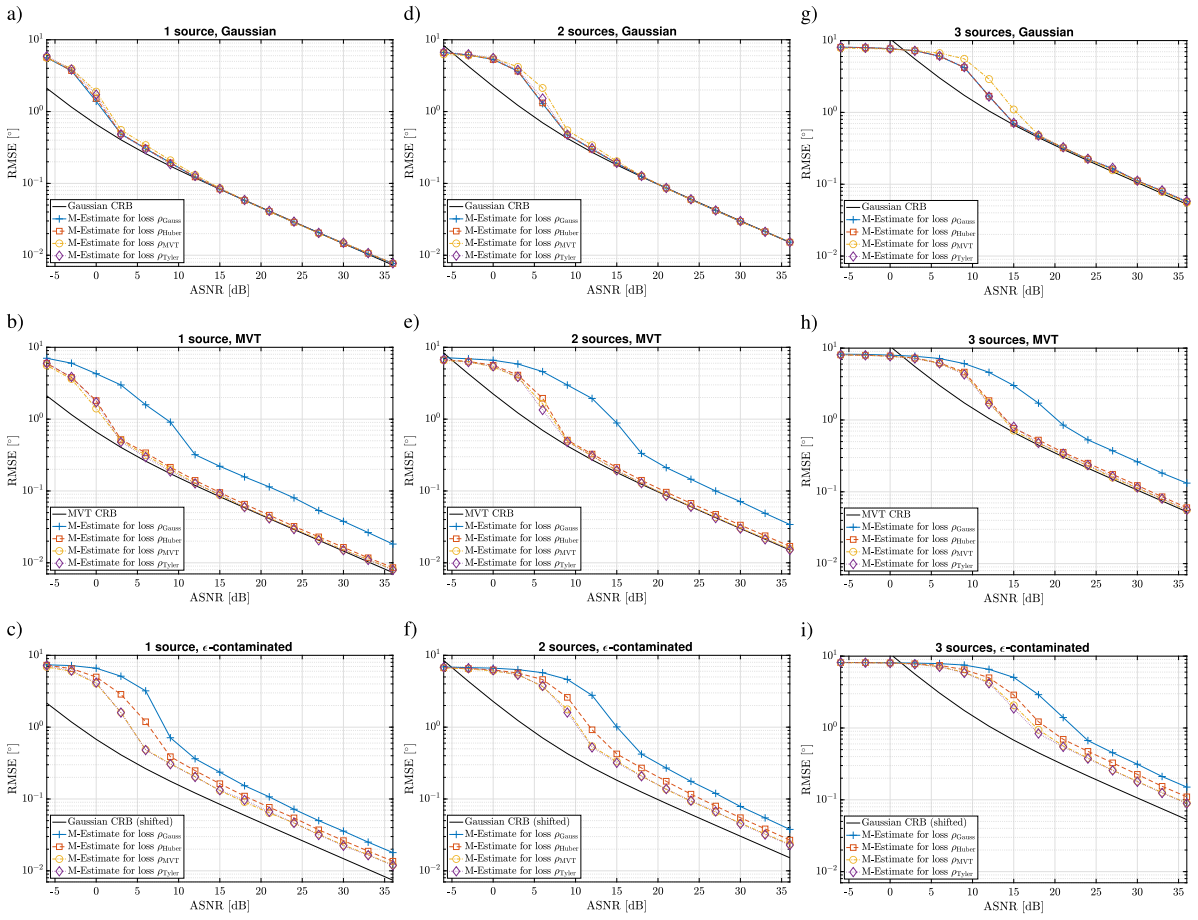

Fig. 1. RMSE of DOA estimators vs. ASNR. Left column: for single source at DOA $-10°$. Center column: for two sources at DOAs $-10°$ and $10°$. Right column: for three sources at DOAs $-3°$, $2°$ and $75°$. All: Simulation for ULA, $N = 20$ sensors, $L = 25$ array snapshots, and dictionary size $M = 18001$ corresponding to DOA resolution $0.01°$, averaged over 250 realizations. Array data models: Top row: Gaussian, middle row: MVT ($\nu_{\text{data}} = 2.1$), bottom row: ϵ-contaminated ($\epsilon = 0.05, \lambda = 10$). The CRB is for Gaussian and MVT array data models, see Appendix B.

3 dB, although a small gap at high ASNR remains. The assumption that ν_{data} is known *a priori* is somewhat unrealistic. However, methods to estimate ν_{data} from data are available, e.g., [55]. Gauss loss exhibits largest RMSE at high ASNR in this case.

Results for ϵ-contaminated noise are shown in Fig. 1(c) for outlier probability $\epsilon = 0.05$ and outlier strength $\lambda = 10$. The resulting noise variance (38) for this heavy-tailed distribution is $\sigma^2 = 5.95\sigma_1^2$. The M-estimators for Tyler loss and MVT-loss ρ_{MVT} perform best followed by the M-estimator for Huber loss ρ_{Huber} with an ASNR penalty of about 2 dB. Worst RMSE exhibits the (non-robust) DOA estimator for Gauss loss ρ_{Gauss} indicating strong impact of outliers on the original (non-robust) SBL algorithm for DOA estimation [11].

Two source scenario. Next, we consider the two-source scenario ($K = 2$) in Table 3. Here, ASNR $= N/\sigma^2$, cf. [54, Eq. (8.112)]. Fig. 1 shows results for RMSE of DOA estimates in scenarios with $L = 25$ snapshots and $N = 20$ sensors.

Simulations for Gaussian array data are shown in Fig. 1(d). Here, the M-estimate for Huber loss ρ_{Huber} for loss parameter $q = 0.9$ performs equally well as the M-estimate for Gauss loss ρ_{Gauss} which is equivalent to the original (non-robust) SBL algorithm for DOA estimation [11]. They approach the CRB for ASNR greater 9 dB. The DOA estimator for Tyler loss performs slightly worse than the previous two. Here, MVT-loss ρ_{MVT} for loss parameter $\nu_{\text{loss}} = 2.1$ has highest RMSE in DOA M-estimates.

Fig. 1(e) shows simulations for heavy-tailed MVT array data with parameter $\nu_{\text{data}} = 2.1$ being small. We observe that M-estimation with Tyler loss and MVT-loss ρ_{MVT} for loss parameter $\nu_{\text{loss}} = 2.1$ perform best, closely followed by M-estimation with Huber loss ρ_{Huber} with loss parameter $q = 0.9$. The non-robust DOA estimator for ρ_{Gauss} performs much worse than the other two showing an ASNR penalty of about 6 dB at high ASNR.

Here, the loss parameter ν_{loss} used by the M-estimator for MVT-loss ρ_{MVT} is identical to the MVT array data parameter ν_{data} used in generating the MVT-distributed array data. In Fig. 1(e), the RMSE for ρ_{MVT} closely follows the MVT CRB (B.1) for ASNR > 9 dB, although a small gap at high ASNR remains.

Results for ϵ-contaminated noise are shown in Fig. 1(f) for outlier probability $\epsilon = 0.05$ and outlier strength $\lambda = 10$. M-estimation with Tyler loss and MVT-loss ρ_{MVT} for loss parameter $\nu_{\text{loss}} = 2.1$ show lowest RMSE followed by Huber loss with slight ASNR penalty. Gauss loss has again the worst performance.

Three source scenario. Array data $\boldsymbol{Y}$ are generated for the three source scenario ($K = 3$), so that $\text{tr}[\boldsymbol{\Gamma}] = 1$. The RMSE performance shown in Figs. 1(g,h,i) are very similar to the corresponding results for the two-source scenario shown in Figs. 1(d,e,f).

Effect of loss function. The effect of the loss function on RMSE performance at high ASNR = 30 dB is illustrated in Fig. 2. This shows that for Gaussian array data all choices of loss functions perform equally well at high ASNR. For MVT data in Fig. 2 (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 array data in Fig. 2 (right) the Gauss loss performs a factor worse than the robust loss functions. Huber loss has slightly higher RMSE than MVT and Tyler loss.

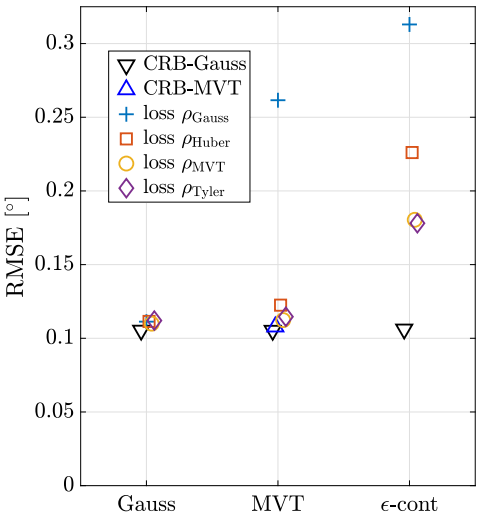

Fig. 2. RMSE for each DOA M-estimator at high ASNR = 30 dB. for each of three array data models (Gaussian, MVT and ϵ-contaminated).

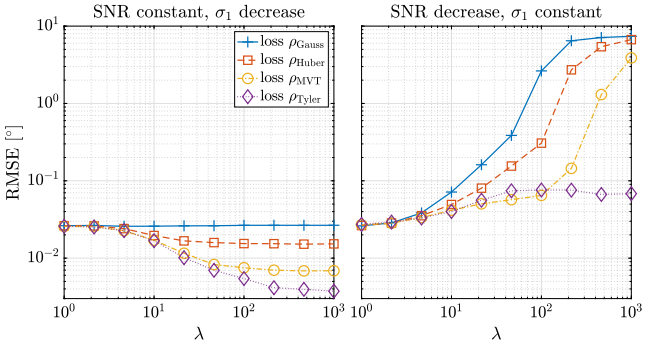

Fig. 3. For ϵ-contaminated array data $\epsilon = 0.05$ with one source, RMSE vs. outlier strength λ for each loss function for (left): ASNR= 25 dB and background noise σ_1 is decreasing, and (right): Background noise σ_1 is fixed and outlier noise $\lambda\sigma_1$ is increasing (at $\lambda = 1$: ASNR= 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.

Effect of outlier strength on RMSE. For small outlier strength λ for ϵ-contaminated data, the Gauss loss performs satisfactorily, but as the outlier noise increases the robust loss functions outperform, see Fig. 3. As λ increases, the total noise changes, see (38). In Fig. 3(left) the total noise is kept constant by decreasing the background noise with increasing λ. In Fig. 3(right) the background noise level is constant whereby the total noise increases. For large noise outlier, Tyler loss clearly has best performance in Fig. 3(left) and does not breakdown in Fig. 3(right).

Comparison with other algorithms. The RMSE of DOA estimates from Table 2 is compared with G-MUSIC with SCM (15) [56], G-MUSIC [57], as well as with Dai and So [12] in Figs. 4 and 5. The Dai and So algorithm [12] performs well for ϵ-contaminated noise, as the outliers are assumed Gaussian.

Coherent sources. Coherent sources are often encountered in applications, e.g., ocean acoustics and radar array processing. In contrast to subspace methods SBL performs well for coherent sources [13,58,59]. To demonstrate this, we report results for a two-source scenario with DOAs at $\pm 10°$ with source correlation coefficient

$$\rho = \frac{\mathsf{E}[x_{8001,l} x^*_{10001,l}]}{\sqrt{\mathsf{E}[|x_{8001,l}|^2]\,\mathsf{E}[|x_{10001,l}|^2]}}. \tag{39}$$

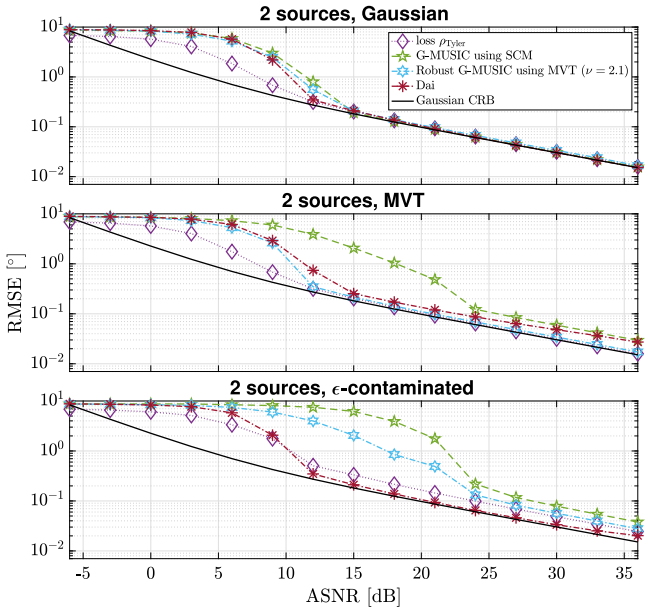

Fig. 4. RMSE vs ASNR for Gaussian, MVT, and ϵ-contaminated array data ($\epsilon = 0.05, \lambda = 10$). Similar setup as Fig. 1 with 2 sources at $\pm 10°$.

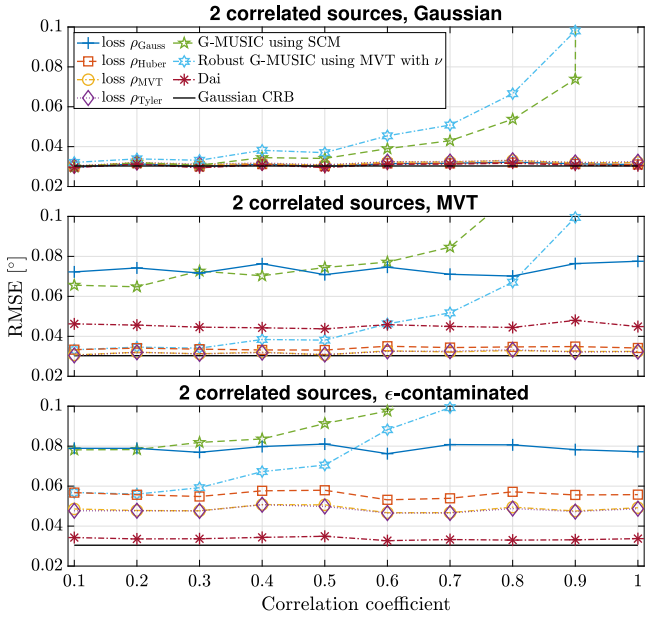

Fig. 5. RMSE vs source correlation for Gaussian, MVT, and ϵ-contaminated array data ($\epsilon = 0.05, \lambda = 10$) at ASNR= 30 dB. Similar setup as Fig. 1 with 2 sources at $\pm 10°$.

The data model for coherent sources has a full source correlation matrix $\boldsymbol{\Gamma}$, but it has been found that estimating a full matrix $\boldsymbol{\Gamma}$ is numerically costly [13,59] and not actually needed for DOA estimation. The Robust SBL algorithm in Table 2 processes coherent sources by ignoring any off-diagonal elements in $\boldsymbol{\Gamma}$, i.e., by using the diagonal $\boldsymbol{\Gamma}$ model (5).

The results in Fig. 5 demonstrate that the coherency of the sources has little effect on the SBL results from Table 2 and this is also true for Dai and So [12], in contrast to G-MUSIC with either SCM (15) [56] or [57] when $\rho > 0.5$.

Effect of dictionary size on RMSE. Due to the algorithmic speedup associated with the DOA grid pruning described in Section 3.5, it is feasible and useful to run the algorithm in Table 2 with large dictionary

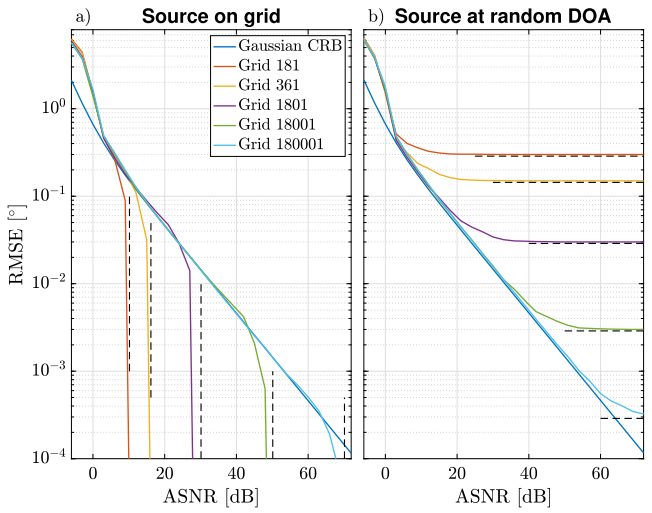

Fig. 6. Effect of dictionary size $M \in \{181, 361, 1801, 18001, 180001\}$ on RMSE vs. ASNR for a single source (a) fixed DOA $-10°$ on the grid and (b) random uniformly distributed DOA $\sim -10° + U(-\delta/2, \delta/2)$. The vertical and horizontal dashed lines indicate asymptotic values due to dictionary size.

size M which translates to the dictionary's angular grid resolution of $\delta = 180°/(M-1)$.

The effect of grid resolution is illustrated in Fig. 6 for a single source impinging on a $N = 20$ element $\lambda/2$-spaced ULA. The Gaussian array data model is used. Fig. 6 shows RMSE vs. ASNR for a dictionary size of $M \in \{181, 361, 1801, 18001, 180001\}$. In Fig. 6(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. 6(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. 6(a), the RMSE performance curve resembles the behavior of an ML-estimator at low ASNR up to a certain threshold ASNR (dashed vertical lines) where the RMSE abruptly crosses the CRB and 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. 6(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 RMSE $= \delta/\sqrt{12}$ for 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. 6(b).

5. Convergence behavior and run time

The DOA M-estimation algorithm in Table 2 uses an iteration to estimate the active set $\mathcal{M}$ whose elements represent the estimated source DOAs. The required number of iterations for convergence of $\mathcal{M}$ depends on the source scenario, array data model, ASNR, and stepsize μ. Under suitable conditions on the parameters, the convergence of the iteration can be shown analytically [60]. We select $\mu = 1$.

Fig. 7 shows the required number of iterations for the three-source scenario versus ASNR and all three array data models. Fig. 7 shows fast convergence for high ASNR, and the number of iterations decreases

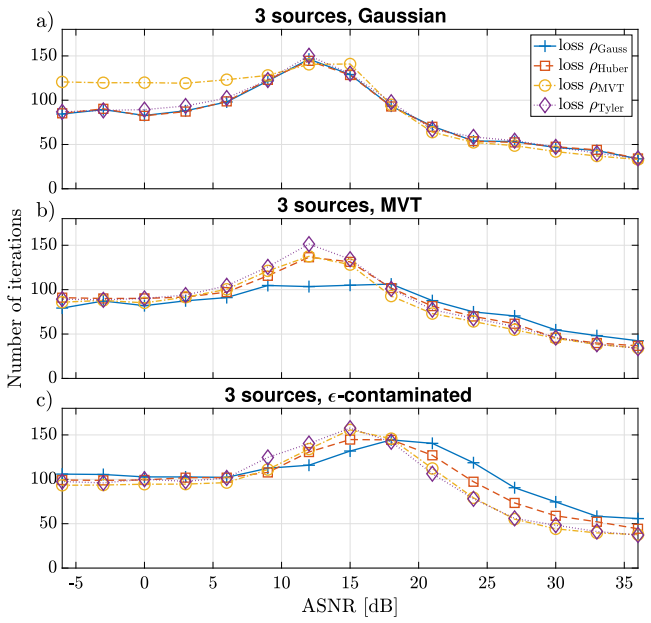

Fig. 7. Iteration count of DOA estimators vs. ASNR for three sources, $N = 20$ sensors, $L = 25$ array snapshots, and dictionary size $M = 18001$ corresponding to DOA resolution 0.01°. Noise: (a) Gaussian, (b) MVT ($\nu_{\text{data}} = 2.1$), (c) ϵ-contaminated ($\epsilon = 0.05, \lambda = 10$).

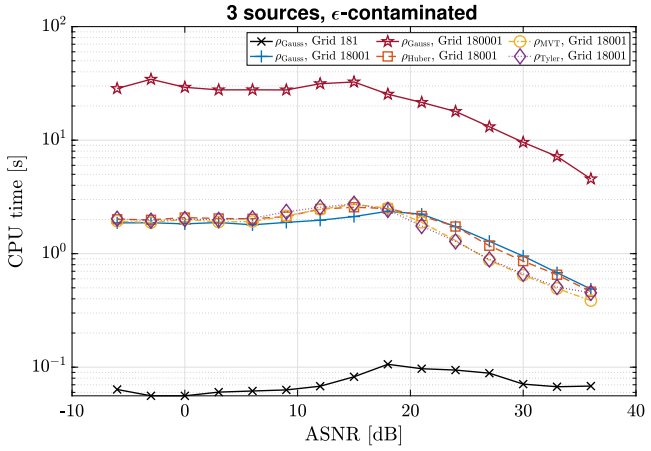

Fig. 8. CPU times for three source scenario vs. ASNR for ϵ-contaminated array data processed with Gauss loss, dictionary size $M \in \{181, 18001, 180001\}$ and for $M = 18001$ for Huber, MVT and Tyler loss.

with increasing ASNR. At ASNR < 5 dB, where the noise dominates, the number of iterations is around 100 and approximately independent of the ASNR. Fig. 7(a) shows that the number of iterations for MVT-loss at low ASNR and Gaussian array data is about 25% larger than for the other loss functions. In the intermediate ASNR range 5–20 dB, the largest number of iterations are required as the algorithm searches the dictionary to find the best matching DOAs. Peak number of iterations is near 160 at ASNR levels between 12 and 15 dB. Fig. 7(c) shows that the number of iterations for Tyler loss and MVT-loss for ϵ-contaminated array data at high ASNR are lowest, followed by Huber and Gauss loss.

The CPU times on an M1 MacBook Pro are shown in Fig. 8 for ϵ-contaminated array data and various choices of dictionary size and loss function. For a fixed dictionary ($M = 18001$), the choice of loss function does not have much influence on the CPU time. At ASNR $>$ 18 dB, MVT and Tyler consume just slightly less CPU time than Huber and Gauss loss. For low ASNR, CPU time increases by a ratio that is approximately proportional to dictionary size 180001/181=1000, but

at high SNR this ratio reduces to 50, due to the efficiency of the DOA grid pruning, cf. Section 3.5. The $M = 180001$ dictionary is quite large in this scenario, but in other scenarios for localization in 3 dimensions, this is an expected dictionary size.

6. Conclusion

Robust and sparse DOA M-estimation is derived based on array data following a zero-mean complex elliptically symmetric distribution with finite second-order moments. Our derivations allowed using different loss functions. The DOA M-estimator is numerically evaluated by iterations and available as a Matlab function on GitHub [51]. A specific choice of loss function determines the RMSE performance and robustness of the DOA estimate. Four choices for loss function are discussed and investigated in simulations with synthetic array data: the ML-loss function for the circular complex multivariate t-distribution with ν degrees of freedom, the loss functions for Huber and Tyler M-estimators. For Gauss loss, the method reduces to Sparse Bayesian Learning. We discuss the robustness of these DOA M-estimators by evaluating the root mean square error for Gaussian, MVT, and ϵ-contaminated array data. The robust and sparse M-estimators for DOA perform well in simulations for MVT and ϵ-contaminated noise and nearly identical with classical SBL for Gaussian noise.

CRediT authorship contribution statement

Christoph F. Mecklenbräuker: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Peter Gerstoft:** Writing – review & editing, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Esa Ollila:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Yongsung Park:** Visualization, Validation, Software, Investigation.

Data availability

No data was used for the research described in the article.

Appendix A. Consistency factor

Here the consistency factor b is evaluated for Huber loss (16), MVT loss, (19), and Tyler loss (21). For elliptical distributions an M-estimator is a consistent estimator of $\alpha\boldsymbol{\Sigma}$, where the constant α is a solution to [3, Eq. (49)]:

$$1 = \mathsf{E}\left[\psi(\boldsymbol{y}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}/\alpha)\right]/N, \tag{A.1}$$

where $\psi(t) = t\,u(t) = t\,\mathrm{d}\rho(t)/\mathrm{d}t$ as defined below (13).

Assuming $\boldsymbol{y} \sim \mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \boldsymbol{\Sigma})$, then we scale the chosen loss function $\rho(t)$ such that (A.1) holds for $\alpha = 1$. Namely, for

$$\rho_b(t) = \rho(t)/b \text{ and } u_b(t) = u(t)/b, \tag{A.2}$$

where b is a scaling constant defined in (12), it clearly holds that $1 = \mathsf{E}\left[\psi_b(\boldsymbol{y}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y})\right]/N$ for $\psi_b(t) = tu_b(t)$. This implies that $\alpha = 1$ and that the M-estimator with loss $\rho_b(\cdot)$ is consistent to the covariance matrix $\boldsymbol{\Sigma}$ when the array data follows the $\mathbb{C}\mathcal{N}_N(\boldsymbol{0}, \boldsymbol{\Sigma})$ distribution.

For Huber loss function (16) the b in (13) can be solved in closed form as [4, Sec. 4.4.2]

$$b = \frac{1}{N}\int_0^\infty (t/2)u_{\text{Huber}}(t/2)f_{\chi^2_{2N}}(t)\mathrm{d}t \tag{A.3}$$

$$= \frac{1}{2N}\int_0^{2c^2} t f_{\chi^2_{2N}}(t)\mathrm{d}t + \frac{1}{N}\int_{2c^2}^\infty c^2 f_{\chi^2_{2N}}(t)\mathrm{d}t \tag{A.4}$$

$$= F_{\chi^2_{2(N+1)}}(2c^2) + c^2(1 - F_{\chi^2_{2N}}(2c^2))/N, \tag{A.5}$$

where $F_{\chi^2_n}(x) = \mathrm{P}\{X \le x\}$ is the cumulative distribution of a central χ^2 distributed random variable X with n degrees of freedom.

For the MVT loss (19) we evaluate (13) by numerical integration.

For Tyler loss $\psi(t) \equiv N\ \forall t$, indicating that the consistency factor for Tyler loss cannot be found based on (A.1). However it is possible to construct an affine equivariant version of Tyler's M-estimator as explained in [61] that estimates the true scatter matrix $\boldsymbol{\Sigma}$. This is accomplished by defining b as the mean of Tyler's weights

$$b_{\text{Tyler}} = \frac{1}{L}\sum_{\ell=1}^{L} u_{\text{Tyler}}(\boldsymbol{y}_\ell\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell). \tag{A.6}$$

Appendix B. CRB for multiple DOA estimation

Gaussian array data. The CRB for multiple DOA estimation is evaluated according to [54, Eq. (8.106)]. In Figures showing RMSE vs. ASNR, the CRB is plotted for $\sigma^2 = N/\text{ASNR}$.

CES array data. The CRB for multiple DOA estimation is derived in [62, Eq. (20)] and [63, Eq. (17)] and [30,31] based on the Slepian–Bangs formula. For DOA parameters, the semi-parametric CRB [28,29] equals the CRB for DOA because the orthogonal projection of the score vector of the DOA parameters on the nuisance tangent space is null (almost surely).

Starting from $p_{\boldsymbol{y}}(\boldsymbol{y}) = p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{\theta})$ given the true source scenario $\boldsymbol{\theta}$ as defined in (2), this gives

$$C_{\text{CR,CES}}(\boldsymbol{\theta}) = \mathrm{tr}\{\boldsymbol{F}^{-1}\}/L, \tag{B.1}$$

where the Fisher information matrix $\boldsymbol{F}$ has elements

$$F_{i,j} = \mathsf{E}\left[\frac{\partial \log p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial \log p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{\theta})}{\partial\theta_j}\right] = (\psi_1 - 1)\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i)\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) + \psi_1\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j) \tag{B.2}$$

with array covariance $\boldsymbol{\Sigma}$ defined in (4), $\boldsymbol{\Sigma}_i = \frac{\partial\boldsymbol{\Sigma}}{\theta_i}$ and

$$\psi_1 = \frac{\mathsf{E}[\psi(\boldsymbol{y}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y})^2]}{N(N+1)}. \tag{B.3}$$

For the MVT array data model, this evaluates to

$$\psi_1^{MVT} = \frac{2N + \nu_{\text{data}}}{2(N+1) + \nu_{\text{data}}} \tag{B.4}$$

and the MVT CRB, $C_{\text{CR,MVT}}(\boldsymbol{\theta})$, is evaluated by (B.1) with (B.2) and (B.4).

References

[1] 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.

[2] P. Gerstoft, P.D. Bromirski, Weather bomb induced seismic signals, Science 353 (6302) (2016) 869–870.

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

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

[5] S. Fortunati, R. Grasso, F. Gini, M.S. Greco, K. LePage, Single snapshot DOA estimation by using compressed sensing, in: EURASIP J. Adv. in Signal Process. 2014, vol. 120, 2014, http://dx.doi.org/10.1186/1687-6180-2014-120.

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

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

[8] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.

[9] D.P. Wipf, S. Nagarajan, Beamforming using the relevance vector machine, in: Proc. 24th Int. Conf. Machine Learning, New York, NY, USA, 2007, pp. 1–4, http://dx.doi.org/10.1145/1273496.1273625.

[10] Z. Yang, L. Xie, C. Zhang, Off-grid direction of arrival estimation using sparse Bayesian inference, IEEE Trans. Signal Process. 61 (1) (2012) 38–43.

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

[12] 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, http://dx.doi.org/10.1109/TSP.2017.2773420.

[13] R.R. Pote, B.D. Rao, Robustness of sparse Bayesian learning in correlated environments, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2020, pp. 9100–9104.

[14] R.R. Pote, B.D. Rao, Maximum likelihood-based gridless DOA estimation using structured covariance matrix recovery and SBL with grid refinement, IEEE Trans. Signal Process. 71 (2023) 802–815.

[15] X. Wu, W.-P. Zhu, J. Yan, Direction of arrival estimation for off-grid signals based on sparse Bayesian learning, IEEE Sens. J. 16 (7) (2015) 2004–2016.

[16] L. Wang, L. Zhao, G. Bi, C. Wan, L. Zhang, H. Zhang, Novel wideband DOA estimation based on sparse Bayesian learning with Dirichlet process priors, IEEE Trans. Signal Process. 64 (2) (2015) 275–289.

[17] P. Chen, Z. Cao, Z. Chen, X. Wang, Off-grid doa estimation using sparse Bayesian learning in MIMO radar with unknown mutual coupling, IEEE Trans. Signal Process. 67 (1) (2018) 208–220.

[18] S. Nannuru, K. Gemba, P. Gerstoft, W. Hodgkiss, C. Mecklenbräuker, Sparse Bayesian learning with multiple dictionaries, Signal Process. 159 (2019) 159–170, http://dx.doi.org/10.1016/j.sigpro.2019.02.003.

[19] J. Yang, Y. Yang, Sparse Bayesian DOA estimation using hierarchical synthesis LASSO priors for off-grid signals, IEEE Trans. Signal Process. 68 (2020) 872–884.

[20] Q. Wang, H. Yu, J. Li, F. Ji, F. Chen, Sparse Bayesian learning using generalized double Pareto prior for DOA estimation, IEEE Signal Process. Lett. 28 (2021) 1744–1748.

[21] 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.

[22] X. Zhang, M.N.E. Korso, M. Pesavento, Maximum likelihood and maximum a posteriori direction-of-arrival estimation in the presence of SIRP noise, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2016, pp. 3081–3085, http://dx.doi.org/10.1109/ICASSP.2016.7472244.

[23] B. Meriaux, X. Zhang, M.N.E. Korso, M. Pesavento, Iterative marginal maximum likelihood DOD and DOA estimation for MIMO radar in the presence of SIRP clutter, Signal Process. 155 (2019) 384–390, http://dx.doi.org/10.1016/j.sigpro.2018.09.034.

[24] J.A. Fishbone, L. Mili, Highly robust complex covariance estimators with applications to sensor array processing, IEEE Open J. Signal Process. 4 (2023) 208–224, http://dx.doi.org/10.1109/OJSP.2023.3261806.

[25] P. Gerstoft, S. Nannuru, C.F. Mecklenbräuker, G. Leus, DOA estimation in heteroscedastic noise, Signal Process. 161 (2019) 63–73, http://dx.doi.org/10.1016/j.sigpro.2019.03.014.

[26] C.D. Richmond, Adaptive Array Signal Processing and Performance Analysis in Non-Gaussian Environments (Ph.D. dissertation), Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., 1996, https://dspace.mit.edu/handle/1721.1/11005.

[27] M.S. Greco, S. Fortunati, F. Gini, Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched conditions, Signal Process. 104 (2014) 381–386.

[28] S. Fortunati, F. Gini, M.S. Greco, A.M. Zoubir, M. Rangaswamy, Semiparametric inference and lower bounds for real elliptically symmetric distributions, IEEE Trans. Signal Process. 67 (1) (2019) 164–177.

[29] S. Fortunati, F. Gini, M.S. Greco, A.M. Zoubir, M. Rangaswamy, Semiparametric CRB and Slepian–Bangs formulas for complex elliptically symmetric distributions, IEEE Trans. Signal Process. 67 (20) (2019) 5352–5364.

[30] H. Abeida, J.-P. Delmas, Slepian-Bangs formula and Cramer Rao bound for circular and non-circular complex elliptical symmetric distributions, IEEE Signal Process. Lett. 26 (10) (2019).

[31] H. Abeida, J.-P. Delmas, Refinement and derivation of statistical resolution limits for circular or rectilinear correlated sources in CES data models 195, 2022.

[32] D.P. Wipf, B. Rao, Sparse Bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164.

[33] Z. Zhang, B.D. Rao, Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning, IEEE J. Sel. Top. Signal Process. 5 (5) (2011) 912–926.

[34] 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 Commun. 11 (10) (2012) 1–11, http://dx.doi.org/10.1109/TWC.2012.090312.111912.

[35] 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.

[36] J. Böhme, Source-parameter estimation by approximate maximum likelihood and nonlinear regression, IEEE J. Oceanic Eng. 10 (3) (1985) 206–212, http://dx.doi.org/10.1109/JOE.1985.1145098.

[37] A.G. Jaffer, Maximum likelihood direction finding of stochastic sources: A separable solution, ICASSP-88, in: IEEE Int. Conf. on Acoust., Speech, and Sig. Process., vol. 5, 1988, pp. 2893–2896, http://dx.doi.org/10.1109/ICASSP.1988.197258.

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

[39] E. Ollila, V. Koivunen, Robust antenna array processing using M-estimators of pseudo-covariance, PIMRC 2003, in: 14th IEEE Proc. Personal, Indoor and Mobile Radio Communications, vol. 3, 2003, pp. 2659–2663, http://dx.doi.org/10.1109/PIMRC.2003.1259213.

[40] C. Mecklenbräuker, P. Gerstoft, E. Ollila, Qualitatively robust bayesian learning for DOA from array data using M-estimation of the scatter matrix, in: 25th International ITG Workshop on Smart Antennas, WSA 2021, Sophia-Antipolis, France, 2021.

[41] C. Mecklenbräuker, P. Gerstoft, E. Ollila, DOA M-estimation using sparse Bayesian learning, in: 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), ICASSP 2022, 2022, pp. 4933–4937, http://dx.doi.org/10.1109/ICASSP43922.2022.9746740.

[42] C. Mecklenbräuker, P. Gerstoft, E. Ollila, Y. Park, Robust sparse Bayesian learning for DOA, in: 31st European Signal Processing Conference, EUSIPCO 2023, 2023.

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

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

[45] M. Mahot, F. Pascal, P. Forster, J.-P. Ovarlez, Asymptotic properties of robust complex covariance matrix estimates, IEEE Trans. Signal Process. 61 (13) (2013) 3348–3356, http://dx.doi.org/10.1109/TSP.2013.2259823.

[46] P.J. Huber, Robust estimation of a location parameter, Ann. Math. Stat. 35 (1) (1964) 73–101.

[47] D.E. Tyler, A distribution-free M-estimator of multivariate scatter, Ann. Statist. 15 (1) (1987) 234–251, URL http://www.jstor.org/stable/2241079.

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

[49] A.H. Sayed, Adaptive Filters, John Wiley & Sons, 2008.

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

[51] Y. Park, E. Ollila, P. Gerstoft, C. Mecklenbräuker, RobustSBL repository, in: GitHub, 2022, https://github.com/NoiseLabUCSD/RobustSBL.

[52] F. Meyer, P. Braca, P. Willett, F. Hlawatsch, A scalable algorithm for tracking an unknown number of targets using multiple sensors, IEEE Trans. Signal Process. 65 (13) (2017) 3478–3493, http://dx.doi.org/10.1109/TSP.2017.2688966.

[53] Z. Zhang, Z. Shi, Y. Gu, Ziv-Zakai bound for DOAs estimation, IEEE Trans. Signal Process. 71 (2023) 136–149, http://dx.doi.org/10.1109/TSP.2022.3229946.

[54] H.L. Van Trees, Optimum Array Processing, Wiley-Interscience, New York, 2002, (Ch. 1–10).

[55] F. Pascal, E. Ollila, D.P. Palomar, Improved estimation of the degree of freedom parameter of multivariate t-distribution, in: 2021 29th European Signal Processing Conference, EUSIPCO, IEEE, 2021, pp. 860–864.

[56] X. Mestre, M.A. Lagunas, Modified subspace algorithms for DoA estimation with large arrays, IEEE Trans. Signal Process. 56 (2) (2008) 598–614, http://dx.doi.org/10.1109/TSP.2007.907884.

[57] R. Couillet, F. Pascal, J.W. Silverstein, Robust estimates of covariance matrices in the large dimensional regime, IEEE Trans. Inf. Theor. 60 (11) (2014) 7269–7278, http://dx.doi.org/10.1109/TIT.2014.2354045.

[58] A. Das, W.S. Hodgkiss, P. Gerstoft, Coherent multipath direction-of-arrival resolution using compressed sensing, IEEE J Oceanic Eng. 42 (2) (2017) 494–505.

[59] C. Mecklenbräuker, P. Gerstoft, G. Leus, Sparse Bayesian learning for doa estimation of correlated sources, in: 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop, SAM, IEEE, 2018, pp. 533–537.

[60] C. Mecklenbräuker, P. Gerstoft, E. Ollila, Y. Park, Convergence analysis of robust and sparse M-estimation of DOA, in: 2024 Information Theory and Applications Workshop, IEEE, 2024.

[61] E. Ollila, D.P. Palomar, F. Pascal, Affine equivariant Tyler's M-estimator applied to tail parameter learning of elliptical distributions, IEEE Signal Process. Lett. (2023).

[62] O. Besson, Y. Abramovich, On the Fisher information matrix for multivariate elliptically contoured distributions, IEEE Signal Process. Lett. 20 (11) (2013) 1130–1133.

[63] M.S. Greco, F. Gini, Cramér-Rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions, IEEE Trans. Signal Process. 61 (24) (2013) 6401–6409.