Mecklenbräuker, Christoph F.; Gerstoft, Peter; Ollila, Esa

Qualitatively Robust Bayesian Learning for DOA from Array Data using M-Estimation of the Scatter Matrix

Published in:
WSA 2021 - 25th International ITG Workshop on Smart Antennas

Published: 01/01/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Mecklenbräuker, C. F., Gerstoft, P., & Ollila, E. (2021). Qualitatively Robust Bayesian Learning for DOA from Array Data using M-Estimation of the Scatter Matrix. In *WSA 2021 - 25th International ITG Workshop on Smart Antennas* (pp. 179-184). VDE Verlag GmbH. https://www.vde-verlag.de/proceedings-en/455686032.html

QUALITATIVELY ROBUST BAYESIAN LEARNING FOR DOA FROM ARRAY DATA USING M-ESTIMATION OF THE SCATTER MATRIX

Christoph F. Mecklenbräuker[1], *Peter Gerstoft*[2], *Esa Ollila*[3]

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

ABSTRACT

The qualitative robustness of direction of arrival estimation using Sparse Bayesian Learning (SBL) is assessed by evaluating the corresponding empirical influence function (EIF). The EIF indicates that SBL is sensitive to deviations from the underlying joint Gaussian assumption on signal and noise.

To improve its qualitative robustness, we modify SBL by plugging-in the sample covariance matrix of the phase-only array data instead of the conventional sample covariance. A qualitatively more robust DOA estimate is derived as maximum likelihood estimate based on the complex multivariate t-distribution as the model-distribution for array data. Finally, we discuss and compare the qualitative robustness of the derived DOA estimators by evaluating the corresponding EIFs.

1. INTRODUCTION

This contribution addresses array processing in the presence of additive outliers in the data and heavy-tailed noise. Stochastic Maximum Likelihood parameter estimation in the context of heavy-tailed models is discussed in [1, 2].

Here, direction of arrival (DOA) estimators are developed for a plane wave observed by a sensor array based on a complex multivariate Student t_ν-distribution data model. A surprisingly robust DOA estimate is derived as the maximum likelihood estimate (MLE) based on this model. [2, Sec. 5.4.2], [3].

2. ARRAY DATA MODEL

We observe narrowband waves on N sensors for L snapshots and the corresponding array data is $\boldsymbol{Y} \in \mathbb{C}^{N\times L}$. The unknown zero-mean complex source amplitudes are the elements of $\boldsymbol{X} \in \mathbb{C}^{M\times L}$ where M is the considered number of hypothetical DOAs on a pre-specified grid. We assume that the complex source amplitudes are independent across sources and snapshots, i.e. x_{ml} and $x_{m'l'}$ are independent. A linear regression model relates the array data $\boldsymbol{Y}$ to the source amplitudes $\boldsymbol{X}$,

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

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

In the presence of few stationary sources, the lth column of $\boldsymbol{X}$ is K-sparse and we assume that the sparsity pattern does not vary across snapshots. We define the active set $\mathcal{M} = \{m \in \mathbb{N} | x_{ml} \neq 0\}$ and $\boldsymbol{A}_{\mathcal{M}} \in \mathbb{C}^{N\times K}$ contains only the K "active" columns of $\boldsymbol{A}$. In our setting, $M \gg N > K$ and (1) is underdetermined.

3. ROBUST BAYESIAN LEARNING

SBL is derived under complex Gaussian assumptions on each element of $\boldsymbol{X}$ and $\boldsymbol{N}$. Direction of arrival (DOA) estimation for plane waves using Sparse Bayesian learning (SBL) is proposed in Table I in Ref. [4] and [5]. A numerically efficient SBL implementation is available on GitHub [6]. SBL provides DOA estimates based on the array data sample $\boldsymbol{Y}$ and the sample covariance matrix ($(\cdot)^H$ denotes Hermitian transpose)

$$\boldsymbol{S}_{\boldsymbol{Y}} = \boldsymbol{Y}\boldsymbol{Y}^H / L \tag{2}$$

3.1. SBL applied to phase-only array data

It is simple and useful to normalize the array data magnitude (keeping only the data's phase) for estimating well separated DOAs in heteroscedastic Gaussian noise [7] and leads to robust DOA estimators [8, 9]. We modify SBL in an ad hoc approach by using the phase-only array data

$$\tilde{\boldsymbol{Y}} = \boldsymbol{Y} \oslash |\boldsymbol{Y}| \tag{3}$$

where $\oslash$ denotes element-wise division and $|\boldsymbol{Y}|$ is the matrix of element-wise magnitudes of $\boldsymbol{Y}$. The phase-only sample covariance matrix [9],

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

3.2. M-estimation based on t-distribution

A qualitatively more robust DOA estimate is derived as maximum likelihood estimate based on the complex multivariate t_ν-distribution as model for array data [2, Sec. 5.4.2]. Since $\boldsymbol{X}$ and $\boldsymbol{N}$ are zero-mean, it follows that the array data $\boldsymbol{Y}$ are zero-mean.

Equations (5)-(6) are based on [3]. Assuming that $\boldsymbol{Y}$ follows a complex multivariate t_ν-distribution with $\nu > 0$ degrees of freedom, the log-likelihood for the scatter matrix $\boldsymbol{\Sigma}$ of the array data is proportional to [3]

$$\mathcal{L}(\boldsymbol{\Sigma}) = -\log\det(\boldsymbol{\Sigma}) - \frac{2N+\nu}{2bL}\sum_{\ell=1}^{L}\log\left(1+\frac{2\boldsymbol{y}_\ell^H\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell}{\nu}\right), \quad (5)$$

where $\boldsymbol{y}_\ell$ is the ℓth array data snapshot, b is a consistency factor defined below, and $\boldsymbol{\Sigma}$ is proportional to the array data covariance matrix when it exists,

$$\mathsf{E}(\boldsymbol{y}_\ell\boldsymbol{y}_\ell^H) = \mathsf{E}(\boldsymbol{S}_{\boldsymbol{Y}}) = \frac{\nu}{\nu-2}b\boldsymbol{\Sigma}, \quad \text{for } \nu > 2, \quad (6)$$

where the expectation is taken over $\boldsymbol{y}_\ell \sim \mathbb{C}t_\nu$. Special cases include the complex Cauchy distribution for $\nu = 1$ and the complex Gaussian in the limit $\nu \to \infty$. Consequently,

$$\lim_{\nu\to\infty}\mathcal{L}(\boldsymbol{\Sigma}) = -\log\det(\boldsymbol{\Sigma}) - \frac{1}{b}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_{\boldsymbol{Y}}\right] \quad (7)$$

which highlights the relation to Stochastic Maximum Likelihood (SML) for Gaussian array data and SBL. We conclude that DOA estimates based on maximizing (5) for large $\nu \to \infty$ converge to SBL estimates.

Above, the term b in (5) is a consistency factor defined as $b = \mathsf{E}[\psi_\nu(\|\boldsymbol{y}\|^2)]/N$, where $\psi_\nu(t) = t\rho'_\nu(t)$ and $\boldsymbol{y}$ is a random vector consisting of independent zero mean and unit variance complex Gaussian random variables ($\boldsymbol{y} \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$). This consistency factor b guarantees that the underlying M-estimator that maximizes (5) is a consistent estimator of the covariance matrix for Gaussian array data.

In the following, we assume that $\boldsymbol{\Sigma} = \mathsf{E}(\boldsymbol{y}_\ell\boldsymbol{y}_\ell^H)$, then we may simply write

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^H + \sigma^2\boldsymbol{I}_N, \quad (8)$$
$$\boldsymbol{\Gamma} = \operatorname{diag}(\boldsymbol{\gamma}) \quad (9)$$

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

$$\boldsymbol{\Sigma}^{-1} = \sigma^{-2}\boldsymbol{I}_N - \sigma^{-2}\boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{A}^H\sigma^{-2}, \quad (10)$$
$$\boldsymbol{\Sigma}_{\boldsymbol{x}} = \left(\sigma^{-2}\boldsymbol{A}^H\boldsymbol{A} + \boldsymbol{\Gamma}^{-1}\right)^{-1}. \quad (11)$$

3.2.1. Estimation of Source Power

Similarly to Ref. [4, Sec. III.D], we regard (5) as a function of $\boldsymbol{\gamma}$ and σ^2 and compute the first order derivative

$$\frac{\partial\mathcal{L}}{\partial\gamma_m} = -\boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m + \frac{2N+\nu}{b\nu L}\sum_{\ell=1}^{L}\frac{\|\boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell\|_2^2}{1+\frac{2\boldsymbol{y}_\ell^H\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell}{\nu}} \quad (12)$$

Note the similarity of (12) with Ref. [4, Eq.(21)]. Setting (12) to zero gives

$$\boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m = \frac{2N+\nu}{b\nu L}\sum_{\ell=v1}^{L}\frac{\boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell\boldsymbol{y}_\ell^H\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m}{1+\frac{2\boldsymbol{y}_\ell^H\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell}{\nu}}, \quad (13)$$
$$= \boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{\boldsymbol{Y}}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m, \quad (14)$$

where the M-estimator of the scatter matrix is [2, Sec. 4.4.1]

$$\boldsymbol{R}_{\boldsymbol{Y}} = \frac{1}{Lb}\sum_{\ell=1}^{L}u_\nu(\boldsymbol{y}_\ell^H\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell)\boldsymbol{y}_\ell\boldsymbol{y}_\ell^H. \quad (15)$$

$$\text{and } u_\nu(t) = \left(\frac{2N+\nu}{\nu}\right)\frac{1}{1+2t/\nu} = \frac{\nu+2N}{\nu+2t} \quad (16)$$

which agrees with $u_\nu(t)$ in [2, Sec. 4.4.1] except for the consistency factor b. Note that $\boldsymbol{R}_{\boldsymbol{Y}}$ can be understood as an adaptively weighted sample covariance matrix [2, Sec. 4.3]. $\boldsymbol{R}_{\boldsymbol{Y}}$ is Fisher consistent for the covariance matrix when $\boldsymbol{Y}$ follows a Gaussian, i.e. $\mathsf{E}[\boldsymbol{R}_{\boldsymbol{Y}}] = \boldsymbol{\Sigma}$ thanks to the consistency factor b [2, Sec. 4.4.1]. We multiply (14) by γ_m and obtain the fixed-point equation

$$\gamma_m = \gamma_m\frac{\boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{\boldsymbol{Y}}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m}{\boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m} \quad \forall m \in \mathcal{M},$$
$$= \gamma_m\frac{\frac{1}{Lb}\sum_{\ell=1}^{L}\left|\boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell\sqrt{u_\nu(\boldsymbol{y}_\ell^H\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell)}\right|^2}{\boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m} \quad (17)$$

which is the basis for an iteration to solve for $\boldsymbol{\gamma}$ numerically.

3.2.2. Estimation of Noise Power

We follow [4, Sec. III.E] and rewrite (14) as

$$\boldsymbol{a}_m^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}-\boldsymbol{R}_{\boldsymbol{Y}})\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m = 0 \quad \forall m \in \mathcal{M},$$
$$\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}-\boldsymbol{R}_{\boldsymbol{Y}})\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}} = \boldsymbol{0}. \quad (18)$$

and insert

$$\boldsymbol{\Sigma} = \boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H + \sigma^2\boldsymbol{I}_N, \quad (19)$$

giving

$$\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H + \sigma^2\boldsymbol{I}_N - \boldsymbol{R}_{\boldsymbol{Y}})\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}} = \boldsymbol{0}. \quad (20)$$

We rearrange the terms, cf. [4, Eq. (25)],

$$\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}} = \boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{R}_{\boldsymbol{Y}} - \sigma^2\boldsymbol{I}_N\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}} \quad (21)$$

input $\boldsymbol{Y} \in \mathbb{C}^{N\times L}$ array data to be analyzed
constant $\boldsymbol{A} \in \mathbb{C}^{N\times M}$ dictionary matrix
constants $\nu, K, j_{\max} = 1200$
initialize $(\sigma^2)^{\text{new}} = 0.1$, $\boldsymbol{\gamma}^{\text{new}} = \mathbf{1}$, $\delta_{\min} = 10^{-3}$, $j = 0$
repeat
    $j = j + 1$, $\boldsymbol{\gamma}^{\text{old}} = \boldsymbol{\gamma}^{\text{new}}$, $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma}^{\text{new}})$
    $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^H + (\sigma^2)^{\text{new}}\boldsymbol{I}_N$ (8)
    $\boldsymbol{R}_{\boldsymbol{Y}} = \frac{1}{Lb}\sum_{\ell=1}^{L} u_\nu(\boldsymbol{y}_\ell^H \boldsymbol{\Sigma}^{-1}\boldsymbol{y}_\ell)\boldsymbol{y}_\ell \boldsymbol{y}_\ell^H$ (15)
    $\gamma_m^{\text{new}} = \gamma_m^{\text{old}}\left(\frac{\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1}\boldsymbol{R}_{\boldsymbol{Y}}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m}{\boldsymbol{a}_m^H \boldsymbol{\Sigma}^{-1}\boldsymbol{a}_m}\right)$ (17)
    $\mathcal{M} = \{m \in \mathbb{N} \mid K \text{ largest peaks in } \boldsymbol{\gamma}^{\text{new}}\}$
    $\boldsymbol{A}_{\mathcal{M}} = [\boldsymbol{a}_{m_1}, \ldots, \boldsymbol{a}_{m_K}]$
    $\boldsymbol{P}_N = \boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}}(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{A}_{\mathcal{M}}^H$ (24)
    $(\sigma^2)^{\text{new}} = \frac{\text{tr}[\boldsymbol{P}_N\boldsymbol{R}_{\boldsymbol{Y}}]}{N-K}$ (28)
    $\delta = \frac{\|\boldsymbol{\gamma}^{\text{new}} - \boldsymbol{\gamma}^{\text{old}}\|_1}{\|\boldsymbol{\gamma}^{\text{old}}\|_1}$
until $\delta \le \delta_{\min}$ or $j > j_{\max}$
output $\mathcal{M}$, $\boldsymbol{\gamma}^{\text{new}}$, $(\sigma^2)^{\text{new}}$

Table 1. Qualitatively Robust Bayesian Learning based on t_νM-estimation of the array data scatter matrix

We now multiply (21) from the left with $\boldsymbol{A}_{\mathcal{M}}(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}})^{-1}$ and with its Hermitian transpose from the right. This gives

$$\boldsymbol{P}\boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{P}^H = \boldsymbol{P}\left(\boldsymbol{R}_{\boldsymbol{Y}} - \sigma^2\boldsymbol{I}_N\right)\boldsymbol{P}^H, \tag{22}$$

where we have defined the projection matrix

$$\boldsymbol{P} = \boldsymbol{A}_{\mathcal{M}}(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1} \tag{23}$$
$$= \boldsymbol{A}_{\mathcal{M}}(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{A}_{\mathcal{M}}^H = \boldsymbol{P}^2 = \boldsymbol{P}^H. \tag{24}$$

A proof for (23) indeed equalling the orthogonal projection matrix (24) onto the signal subspace is given in the Appendix. We use $\boldsymbol{P}\boldsymbol{A}_{\mathcal{M}} = \boldsymbol{A}_{\mathcal{M}}$ and simplify (22) to

$$\boldsymbol{A}_{\mathcal{M}}\boldsymbol{\Gamma}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^H = \boldsymbol{P}\left(\boldsymbol{R}_{\boldsymbol{Y}} - \sigma^2\boldsymbol{I}_N\right)\boldsymbol{P}^H \tag{25}$$
$$\boldsymbol{\Sigma} - \sigma^2\boldsymbol{I}_N = \boldsymbol{P}\left(\boldsymbol{R}_{\boldsymbol{Y}} - \sigma^2\boldsymbol{I}_N\right)\boldsymbol{P}^H \tag{26}$$
$$\sigma^2(\boldsymbol{I}_N - \boldsymbol{P}\boldsymbol{P}^H) = \left(\boldsymbol{\Sigma} - \boldsymbol{P}\boldsymbol{R}_{\boldsymbol{Y}}\boldsymbol{P}^H\right) \tag{27}$$

which resembles [4, Eq. (26)].

Finally, we evaluate the trace,

$$\sigma^2 = \frac{\text{tr}\left[\boldsymbol{P}_N\boldsymbol{R}_{\boldsymbol{Y}}\right] + \epsilon}{N-K} \approx \frac{\text{tr}\left[\boldsymbol{P}_N\boldsymbol{R}_{\boldsymbol{Y}}\right]}{N-K} \tag{28}$$

where $\boldsymbol{P}_N = \boldsymbol{I}_N - \boldsymbol{P}$ is the projection to the noise subspace and $\epsilon = \text{tr}\left[\boldsymbol{\Sigma} - \boldsymbol{R}_{\boldsymbol{Y}}\right]$, cf. [4, Eq. (27)]. The error ϵ is zero mean because $E[\boldsymbol{R}_{\boldsymbol{Y}}] = \boldsymbol{\Sigma}$.

4. SIMULATION RESULTS

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

For the results a single plane wave with DOA $-45°$ with additive noise is observed by a uniform linear antenna array with $N = 20$ elements at half-wavelength spacing. The transfer matrix $\boldsymbol{A}$ has $M = 181$ columns and the steering vectors are computed for the angular grid $\{0°, 1°, \ldots, 179°\}$. Three types of zero-mean circularly symmetric complex-valued noise $\boldsymbol{N} = [\boldsymbol{n}_1 \ldots \boldsymbol{n}_L]$ in (1) are simulated i.i.d:

Gaussian: $\boldsymbol{n}_\ell \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I}_N)$. This is the nominal noise distribution and this is the standard assumption in array processing.

Complex Student: $\boldsymbol{n}_\ell \sim \mathbb{C}t_\nu$-distributed. The noise covariance matrix is $\sigma^2\boldsymbol{I}_N$, cf. [3] and [2, Sec. 4.2.2]. The limiting distribution of $\mathbb{C}t_\nu$-distributed noise for $\nu \to \infty$ is Gaussian.

ϵ-contaminated: $\boldsymbol{n}_\ell$ is drawn from $\mathbb{C}\mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I}_N)$ with probability $(1-\epsilon)$ and with probability ϵ from $\mathbb{C}\mathcal{N}(\mathbf{0}, (\lambda\sigma)^2\boldsymbol{I}_N)$, The resulting noise covariance matrix is $(1 - \epsilon + \epsilon\lambda^2)\sigma^2\boldsymbol{I}_N$. The limiting distribution of ϵ-contaminated noise for $\epsilon \to 0$ and any choice of $\lambda = \text{const.} > 0$ is Gaussian.

t_νM-estimation for DOA using the log-likelihood (5) for $\mathbb{C}t_\nu$-distributed array data is implemented via the algorithm documented in Table 1. For performance comparison, we evaluate DOA estimates by the conventional beamformer (CBF), by the SBL implementation `SBL_v4.m` [6] and by SBL applied to phase-only data (3) as described in Sec. 3.1 for identical synthetic data realizations $\boldsymbol{Y}$. The Cramér-Rao Bound (CRB) for DOA estimation for a single source in additive white Gaussian noise (AWGN) is also evaluated [10, Eq. (8.130)].

Figure 1 shows obtained results for Root Mean Squared Error of DOA estimates in scenarios with $L = 25$ snapshots and $N = 20$ sensors. RMSE is averaged over $4 \cdot 10^4$ i.i.d. realizations of DOA estimates from array data $\boldsymbol{Y}$. In these scenarios, we have more snapshots than sensors, $L > N$, which ensures that $\boldsymbol{S}_{\boldsymbol{Y}}$, $\tilde{\boldsymbol{S}}_{\tilde{\boldsymbol{Y}}}$, and $\boldsymbol{R}_{\boldsymbol{Y}}$ have full rank almost surely.

Simulation results for Gaussian noise are shown in Fig. 1(a). In this scenario, the CBF implements the maximum-likelihood DOA estimator and the CBF approaches the CRB for SNR greater 5 dB. All three SBL-type DOA estimators perform very similarly and slightly worse than the CBF. SBL-type algorithms for DOA are biased and their RMSE may be lower than the CRB. The DOA bias for SBL-type algorithms depends on the dictionary size (here: $M = 181$ which corresponds $1°$ angular resolution). We note that SBL4 and SBL4 t_νM-estimator with parameter $\nu = 2.1$ perform identically. This is due to the consistency factor b introduced in (5).

Figure 1(b) shows simulation results when the noise follows a $\mathbb{C}t_\nu$-distribution with $\nu = 2.1$ being small. We observe that the t_νM-estimator for DOA (Table 1) performs best, closely followed by phase-only processing with SBL4. Here, the ν parameter of the t_νM-estimator is chosen identical to the ν of the noise model. The assumption that the value of ν is perfectly known is somewhat unrealistic. We note that the original SBL4 algorithm and the CBF exhibit a large gap in SNR compared to the previous two DOA estimators.

The results for ϵ-contaminated noise are shown in Fig. 1(c) for $\epsilon = 0.05$ and $\lambda = 10$ and the noise variance evaluates to $5.95\sigma^2$. In this scenario, the phase-only processing with SBL4 shows lowest RMSE in its DOA estimates, followed by the t_νM-estimator. Poorest RMSE performance have CBF and SBL4 which shows that outliers in the array data severely impact CBF and SBL4.

5. QUALITATIVE ROBUSTNESS

Qualitative robustness of a DOA estimator T is measured with the influence function (IF), cf. [9, 2, 11, 12]. The IF measures the sensitivity of T to a small change in the array data distribution in the neighborhood of the assumed array data distribution F_0,

$$\mathrm{IF}(\boldsymbol{z};\, \mathrm{T}, F_0) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \Big(\mathrm{T}(F_\varepsilon) \overset{\mathrm{a}}{-} \mathrm{T}(F_0) \Big) \tag{29}$$

where $F_\varepsilon = (1-\varepsilon)F_0 + \varepsilon G$ is the ε-contaminated array data distribution and $\theta_1 \overset{\mathrm{a}}{-} \theta_2$ denotes the *angular difference*

$$\theta_1 \overset{\mathrm{a}}{-} \theta_2 = \min\big(|\theta_1 - \theta_2| \,,\ 360^\circ - |\theta_1 - \theta_2| \big) \tag{30}$$

which is valid for $|\theta_1 - \theta_2| \le 360^\circ$. The ε-contaminated distribution F_ε is the nominal distribution F_0 with an additive outlier $\boldsymbol{z} \in \mathbb{C}^N$ in the array data. Thus, (29) measures the DOA error due to an infinitesimal additive contamination $\boldsymbol{z}$ on the DOA estimator T, standardized by the mass of the contamination. A qualitatively robust estimator is characterized by an IF that is continuous and bounded.

Let $\delta_{\boldsymbol{z}}$ be the complex point-mass distribution function at $\boldsymbol{z} \in \mathbb{C}^N$. Hence $\delta_{\boldsymbol{z}}$ is chosen as a random point on the N-dimensional real-valued unit-hypersphere scaled by a complex weight $re^{j\phi}$, where ϕ is uniformly distributed in $[0,\, 2\pi]$. The magnitude r is denoted as outlier radius in the sequel. All elements of the outlier $\boldsymbol{z}$ have the same phase angle which is similar to a plane wave arriving from broadside at a uniform linear array. Instead of evaluating the IF analytically, we use a consistent estimator for IF, namely the empirical influence function (EIF) [12].

The EIF for DOA estimates is here defined as[1]

$$\begin{aligned} \mathrm{EIF}(\boldsymbol{z};\, \hat{\theta}(\cdot), \boldsymbol{Y}) &= \mathbb{E}_{\delta_{\boldsymbol{z}}} \left\{ \frac{\hat{\theta}\big(\boldsymbol{Y}_{1:L}\big) \overset{\mathrm{a}}{-} \hat{\theta}\big(\boldsymbol{Y}_{1:L-1}\big)}{\frac{1}{L}} \right\} \\ &= L\, \mathbb{E}_{\delta_{\boldsymbol{z}}} \left\{ \hat{\theta}\big(\boldsymbol{Y}_{1:L}\big) \overset{\mathrm{a}}{-} \hat{\theta}\big(\boldsymbol{Y}_{1:L-1}\big) \right\}, \end{aligned} \tag{31}$$

[1]Notation: $\boldsymbol{Y}_{k:l} = [\boldsymbol{y}_k \dots \boldsymbol{y}_l]$.

where $\hat{\theta}(\cdot)$ is a DOA estimator of interest. Multiple measurement vectors are gathered in the matrix $\boldsymbol{Y}$, where the last column of $\boldsymbol{Y}$, i.e., the Lth array data snapshot $\boldsymbol{y}_L$, contains a contaminated observation. For the numerical evaluation of (31), we have modelled the contamination as a noise outlier affecting the Lth array data snapshot. The array data model (1) is augmented for the additive outlier which is generated as $w \sim \mathbb{C}\mathcal{N}(0, (\lambda\sigma)^2)$ for a selected outlier strength λ,

$$\boldsymbol{Y}_{1:L-1} = \boldsymbol{A}\boldsymbol{X}_{1:L-1} + \boldsymbol{N}_{1:L-1} \tag{32}$$

$$\boldsymbol{y}_L = \boldsymbol{A}\boldsymbol{x}_L + \boldsymbol{n}_L + w\boldsymbol{z}, \tag{33}$$

for some choice of outlier vector $\boldsymbol{z}$ explained below. The EIF according to (31) is the expected angular difference between the DOA estimate for the contaminated sample $\boldsymbol{Y}_{1:L}$ and the DOA estimate for the sample from the nominal distribution $\boldsymbol{Y}_{1:L-1}$. We have evaluated four different models for the outliers in Figure 2a–d which show the computed EIF versus outlier strength λ. The 2-norm of the outlier vector $\boldsymbol{z}$ equals those of the steering vectors in the dictionary $\boldsymbol{A}$. Figure 2a shows the EIF when the outlier vector is constant. The outlier appears in the data of sensor 1, i.e. $\boldsymbol{z} = \sqrt{N}\boldsymbol{e}_1$. Figure 2b shows the EIF when the outlier vector is random. Here $\boldsymbol{z}$ is uniformly distributed on a sphere in $\mathbb{C}^N$ with radius $\sqrt{N}$. Figure 2c shows the EIF when the outlier vector is constant equalling the steering vector for a broadside source, i.e. $\boldsymbol{z} = (1\ 1 \dots 1)^T$. Figure 2d shows the EIF when the outlier vector is a randomly selected steering vector from the dictionary $\boldsymbol{A}$. All steering vectors being equally probable. It is seen that the $t_\nu M$-estimator and phase-only SBL estimates are qualitatively robust because the EIF is much smaller than 0.1° for all λ, in contrast to the EIF of CBF and SBL.

6. CONCLUSION

The EIF indicates that SBL and CBF are sensitive to deviations from the Gaussian distribution of the noise. A qualitatively robust DOA estimate is derived as MLE for complex multivariate t_ν-distributed array data which performs well for Gaussian and ϵ-contaminated noise. We also modify SBL by plugging-in the phase-only sample covariance. Finally, we discuss the DOA performance and qualitative robustness of these DOA estimators by evaluating the corresponding EIFs.

7. APPENDIX

Here, we prove that $\boldsymbol{P}$ defined in (23) is the orthogonal projection matrix (24) onto the signal subspace, $\mathrm{span}(\boldsymbol{A}_{\mathcal{M}})$.

Note that the $N-K$ smallest eigenvalues of the $N \times N$ array covariance matrix $\mathsf{E}(\boldsymbol{S}_{\boldsymbol{Y}})$ are σ^2 and the corresponding eigenvectors are orthogonal to the columns of $\boldsymbol{A}_{\mathcal{M}}$. These eigenvectors span the noise subspace and the eigenvectors corresponding to the K largest eigenvalues span the signal subspace (**reference here**). Let $\boldsymbol{E} = \begin{bmatrix}\boldsymbol{E}_S & \boldsymbol{E}_N\end{bmatrix} \in \mathbb{C}^{N\times N}$

denote the matrix of eigenvectors of $\mathsf{E}(\boldsymbol{S_Y})$. Thus $\boldsymbol{E}_S \in \mathbb{C}^{N\times K}$ is the matrix of the signal subspace eigenvectors and $\boldsymbol{E}_N \in \mathbb{C}^{N\times(N-K}$ is the matrix of the noise subspace eigenvectors. The orthogonal projection matrix to the signal subspace is $\boldsymbol{\Pi}_S = \boldsymbol{E}_S\boldsymbol{E}_S^H$ and the orthogonal projection matrix to the noise subspace is $\boldsymbol{\Pi}_N = \boldsymbol{I}_N - \boldsymbol{\Pi}_S$.

The eigenvalues of $\mathsf{E}(\boldsymbol{S_Y})$ are denoted by $\lambda_1 \geq \cdots \geq \lambda_K > \lambda_{K+1} = \cdots = \lambda_N = \sigma^2$. The diagonal matrix of eigenvalues $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1,\ldots,\lambda_N)$ is expressed as a block diagonal matrix $\boldsymbol{\Lambda} = \operatorname{blkdiag}(\boldsymbol{\Lambda}_S,\boldsymbol{\Lambda}_N)$, where $\boldsymbol{\Lambda}_S = \operatorname{diag}(\lambda_1,\ldots,\lambda_K)$ consists of signal subspace eigenvalues and $\boldsymbol{\Lambda}_N = \sigma^2\boldsymbol{I}_{N-K}$ consists of noise subspace eigenvalues. Then using the eigenvalue decomposition

$$\mathsf{E}(\boldsymbol{S_Y}) = \begin{bmatrix}\boldsymbol{E}_S & \boldsymbol{E}_N\end{bmatrix}\operatorname{blkdiag}(\boldsymbol{\Lambda}_S,\boldsymbol{\Lambda}_N)\begin{bmatrix}\boldsymbol{E}_S & \boldsymbol{E}_N\end{bmatrix}^H \quad (34)$$

and the fact that $\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{E}_N = \boldsymbol{0}$ and $\boldsymbol{\Sigma} = \tau^{-1}\mathsf{E}(\boldsymbol{S_Y})$ for $\tau = \nu/(\nu-2)$, we obtain $\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1} = \tau(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{E}_S)\boldsymbol{\Lambda}_S^{-1}\boldsymbol{E}_S^H$ and

$$\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}} = \tau(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{E}_S)\boldsymbol{\Lambda}_S^{-1}(\boldsymbol{E}_S^H\boldsymbol{A}_{\mathcal{M}}), \quad (35)$$

Thus $(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}})^{-1} = \tau^{-1}(\boldsymbol{E}_S^H\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{\Lambda}_S(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{E}_S)^{-1}$, and

$$\begin{aligned}\boldsymbol{P} &= \boldsymbol{A}_{\mathcal{M}}(\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{A}_{\mathcal{M}}^H\boldsymbol{\Sigma}^{-1}\\ &= \boldsymbol{A}_{\mathcal{M}}(\boldsymbol{E}_S^H\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{E}_S^H.\end{aligned} \quad (36)$$

Since $\boldsymbol{A}_{\mathcal{M}}$ spans the same subspace as $\boldsymbol{E}_S$, then $\boldsymbol{A}_{\mathcal{M}} = \boldsymbol{E}_S\boldsymbol{L}$ for some non-singular matrix $\boldsymbol{L} \in \mathbb{C}^{K\times K}$. This implies that

$$\begin{aligned}\boldsymbol{P} &= \boldsymbol{A}_{\mathcal{M}}(\boldsymbol{E}_S^H\boldsymbol{A}_{\mathcal{M}})^{-1}\boldsymbol{E}_S^H\\ &= \boldsymbol{E}_S\boldsymbol{L}(\boldsymbol{E}_S^H\boldsymbol{E}_S\boldsymbol{L})^{-1}\boldsymbol{E}_S^H = \boldsymbol{E}_S\boldsymbol{E}_S^H = \boldsymbol{\Pi}_S,\end{aligned} \quad (37)$$

where we used that $\boldsymbol{E}_S^H\boldsymbol{E}_S = \boldsymbol{I}_K$ and that $\boldsymbol{L}$ is non-singular. Since $\boldsymbol{P}$ equals the orthogonal projection matrix $\boldsymbol{\Pi}_S$, it verifies $\boldsymbol{P} = \boldsymbol{P}^2 = \boldsymbol{P}^H$.

8. REFERENCES

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

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

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

[4] 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.

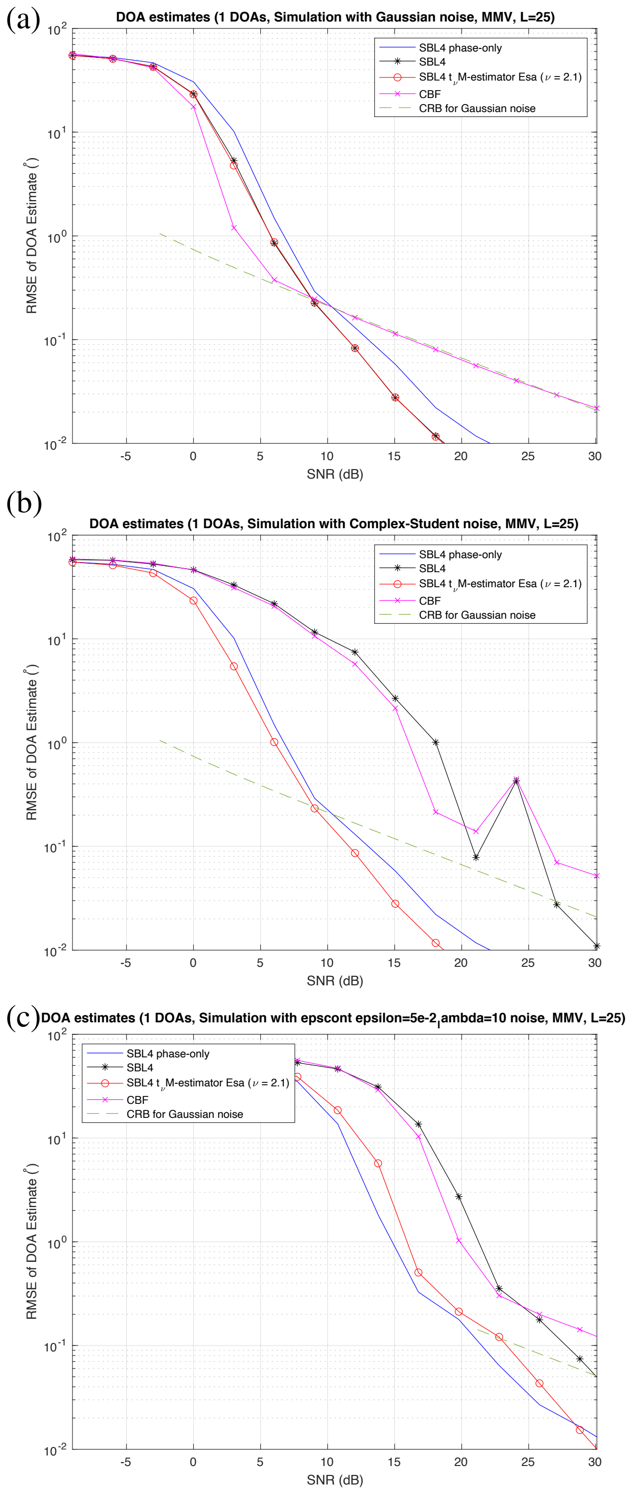

Fig. 1. Root Mean Squared Error (RMSE) of Direction of Arrival (DOA) estimators versus Signal to Noise Ratio (SNR). Simulation with synthetic data for uniform line array, $N = 20$ sensors and $L = 25$ array data snapshots. Noise models: (a) Gaussian, (b) Complex Student t_ν with $\nu = 2.1$, (c) ϵ-contaminated Gaussian with $\epsilon = 0.05, \lambda = 10$. Results averaged over $10^6/L = 4\cdot 10^4$ realizations.

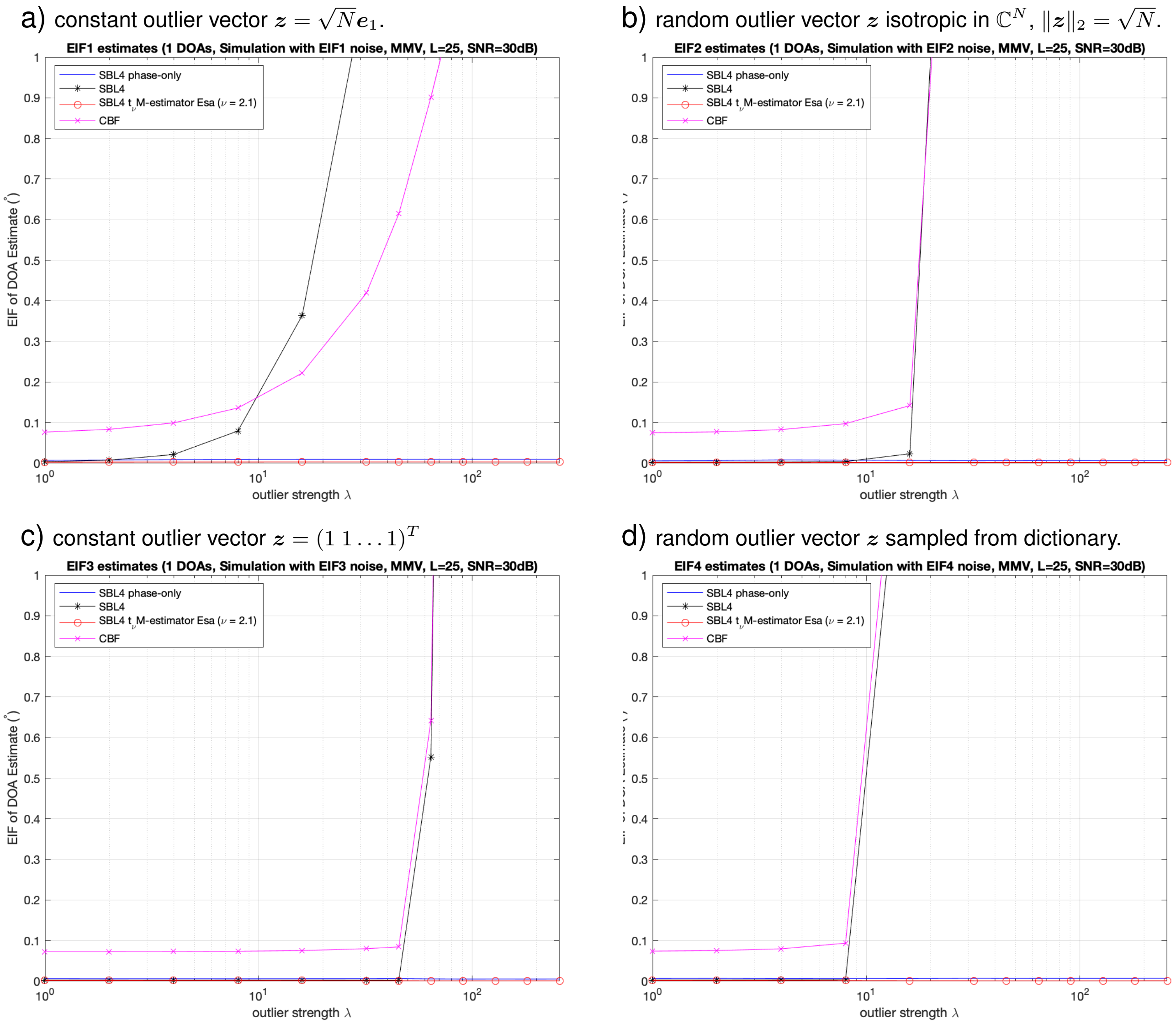

Fig. 2. EIF of DOA estimators for several outlier vectors vs. outlier strength λ. DOA estimators use $L = 25$ snapshots of synthetic data for uniform line array, $N = 20$ sensors, Signal to Noise Ratio 30 dB, averaged over $10^5/L = 4 \cdot 10^3$ realizations.

[5] S. Nannuru, K.L Gemba, P. Gerstoft, W.S. Hodgkiss, and C.F. Mecklenbräuker. Sparse Bayesian learning with multiple dictionaries. *Signal Processing*, 159:159–170, 2019.

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

[7] P. Gerstoft, S. Nannuru, C. F. Mecklenbräuker, and G. Leus. DOA estimation in heteroscedastic noise. *Signal Processing*, 161:63–73, 2019.

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

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

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

[11] E. Ollila and V. Koivunen. Influence functions for array covariance matrix estimators. In *IEEE Workshop on Statistical Signal Processing, 2003*, pages 462–465, Sep. 2003.

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