Convergence Analysis of Robust and Sparse M-Estimation of DOA

Christoph Mecklenbräuker[1], *Peter Gerstoft*[2], *Esa Ollila*[3], *Yongsung Park*[2]

[1]Inst. of Telecommunications, TU Wien, Austria
[2]NoiseLab, UCSD, San Diego, USA
[3]Dept. of Information and Communications Engineering, Aalto Univ., Finland

ITA 2024, San Diego, Feb. 19, 2024

Technische Universität Wien

Aalto University

Contents

Introduction

Array Data Model

M-Estimation based on Complex Elliptically Symmetric (CES)-distribution

Sketch of the Proof

Simulation Results

Conclusion

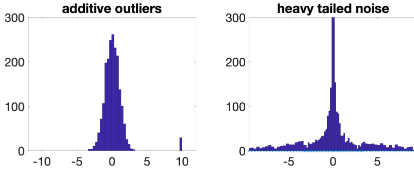
- In this talk, the CES distribution is mentioned a lot
- The CES distribution is defined by the form of its pdf

$$p(\boldsymbol{y}|0, \Sigma) = \det(\Sigma^{-1}) g(\boldsymbol{y}^{\mathrm{H}} \Sigma^{-1} \boldsymbol{y})$$

for a suitable density generator g.

Introduction 2/2

We are interested in array processing for data with
- additive outliers
- heavy-tailed noise



Array data in simulations is one of
- Complex Gaussian
- Multivariate Complex t_ν (MVT)
- ϵ-contaminated Complex Gaussian

M-Estimator for Direction of Arrival (DOA) derived for general loss function

Sparse Bayesian Learning (SBL) approach for this model

ICASSP 2022 contribution generalized IEEE SPL 2016 paper

Contents

Introduction

Array Data Model

M-Estimation based on CES-distribution

Sketch of the Proof

Simulation Results

Conclusion

Array Data Model 1/2 (CES model)

- We observe narrowband waves on N sensors for L snapshots
  - Any array geometry works, but our results are for Uniform Linear Array (ULA) with half wavelength spacing
- $\boldsymbol{y}_\ell$ is the ℓth array data snapshot
- Array data $\boldsymbol{Y} = [\boldsymbol{y}_1 \ldots \boldsymbol{y}_L] \in \mathbb{C}^{N \times L}$ (given)
- Array data covariance matrix $\mathrm{E}(\boldsymbol{y}_\ell \boldsymbol{y}_\ell^{\mathsf{H}})$ exists

Array Data Model 2/2 (CES model)

- We assume $\boldsymbol{y}_\ell \sim$ CES-distributed with unknown scatter matrix Σ

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

for a suitable density generator g.

- Σ is the so-called scatter matrix
- Σ equals the array data covariance matrix $\mathrm{E}(\boldsymbol{y}_\ell \boldsymbol{y}_\ell^H)$ for Gaussian data

Contents

Introduction

Array Data Model

M-Estimation based on CES-distribution

Sketch of the Proof

Simulation Results

Conclusion

M-Estimation based on CES-distribution 1/7

- General approach based on loss functions:
- M-estimator of Σ is minimizer of

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

  where $\rho(\cdot)$ is a chosen loss function

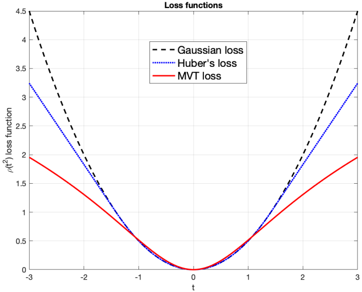
- By introducing the consistency factor b, the maximizer $\hat{\Sigma}$ of $\mathcal{L}(\Sigma)$ becomes consistent for the array data covariance matrix for Gaussian data
  - b enables identifying covariance with scatter matrix, $\mathrm{E}(\boldsymbol{y}_\ell \boldsymbol{y}_\ell^{\mathrm{H}}) = \Sigma$
  - b is evaluated a priori

- Example loss functions:

$$\text{Gaussian loss,} \quad \rho_{\mathrm{G}}(t) = t$$

$$\text{Huber's loss,} \quad \rho_{\mathrm{H}}(t) = \begin{cases} t & \text{for } t \leqslant c^2, \\ c^2\big(\log(t/c^2) + 1\big) & \text{for } t > c^2, \end{cases}$$

$$\text{MVT loss,} \quad \rho_{\mathrm{T}}(t) = \frac{\nu+2N}{2}\log(\nu + 2t)$$



the ρ differ in their weighting of outliers

M-Estimation based on CES-distribution 3/7 (Table)

loss name	data density generator $g(t)$	loss function $\rho(t)$	weight $u(t)$
	$g(t)$	$-\log g(t)$	$\rho'(t)$
Gauss	e^{-t}	t	1
MVT	$(1 + t/\nu)^{-(\nu+2N)/2}$	$\frac{\nu+2N}{2} \log(\nu + 2t)$	$\frac{\nu+2N}{\nu+2t}$
Huber	$\mathrm{e}^{-\rho_{\mathrm{Huber}}(t)}$	$= \begin{cases} t \text{ if } & t \leqslant c^2, \\ c^2\big(\log(t/c^2) + 1\big) & \text{else} \end{cases}$	$= \begin{cases} 1 \text{ if } t < c^2 \\ c^2/t \text{ else} \end{cases}$
Tyler	t^{-N}	$N \log t$	N/t

M-Estimation based on CES-distribution 4/7 (generalized Jaffer's condition)

- $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^{\mathsf{H}} + \sigma^2 \boldsymbol{I}_N$ where $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$ and $\boldsymbol{\gamma} = [\gamma_1 \ldots \gamma_M]^{\mathsf{T}}$
- First order condition: $\frac{\partial \mathcal{L}}{\partial \gamma_m} = 0$
- Generalizes Jaffer's condition (ICASSP 1988) on $\boldsymbol{\Sigma}$

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

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

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

Weight function $u(\cdot) = \rho'(\cdot)$ depends on chosen loss $\rho(\cdot)$

M-Estimation based on CES-distribution 5/7 (Iterating $\boldsymbol{R_Y}$ and Σ)

$$\Gamma^{(j-1)} = \mathrm{diag}(\gamma^{(j-1)})$$

$$\Sigma^{(j-1)} = \boldsymbol{A}\Gamma^{(j-1)}\boldsymbol{A}^H + (\sigma^2)^{(j-1)}\boldsymbol{I}_N$$

$$\boldsymbol{R_Y}^{(j)} = \frac{1}{Lb}\sum_{\ell=1}^{L} u(\boldsymbol{y}_\ell^H(\Sigma^{(j-1)})^{-1}\boldsymbol{y}_\ell;\cdot)\boldsymbol{y}_\ell\boldsymbol{y}_\ell^H$$

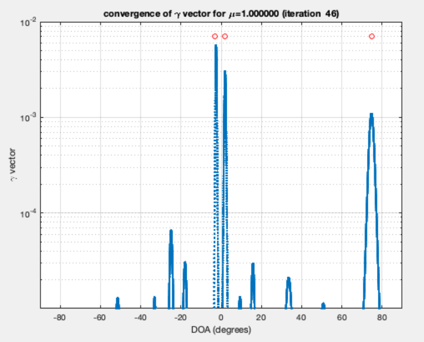
$$\boxed{\gamma_m^{(j)} = (1-\mu)\gamma_m^{(j-1)} + \mu\gamma_m^{(j-1)}\left(\frac{\boldsymbol{a}_m^H\Sigma^{-1}\boldsymbol{R_Y}\Sigma^{-1}\boldsymbol{a}_m}{\boldsymbol{a}_m^H\Sigma^{-1}\boldsymbol{a}_m}\right)}$$

$$j = j+1$$

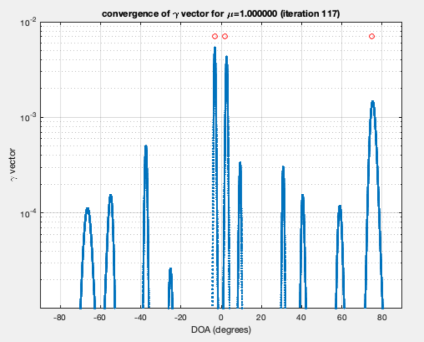
Show animation of iterations: γ vector convergence

Matlab-code

M-Estimation based on CES-distribution 6/7 (Animation for $\mu = 1$)

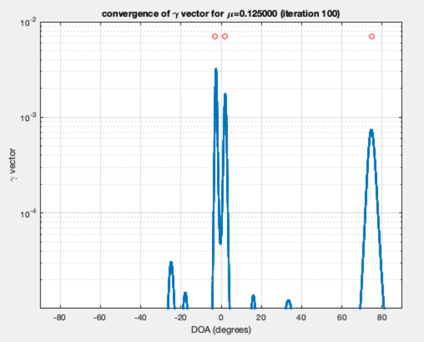


3 sources, $\mu = 1.000$, Gaussian data

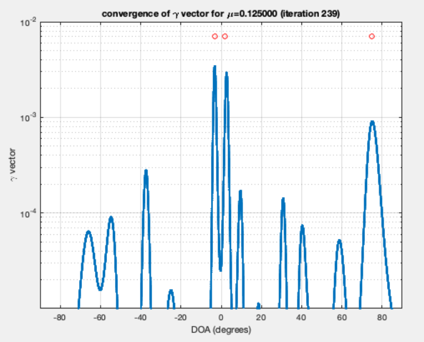


3 sources, $\mu = 1.000$, ϵ-contaminated data

M-Estimation based on CES-distribution 6/7 (Animation for $\mu = 0.125$)



3 sources, $\mu = 0.125$, Gaussian data



3 sources, $\mu = 0.125$, ϵ-contaminated data

$$\boxed{\gamma_m^{(j)} = (1-\mu)\gamma_m^{(j-1)} + \mu\gamma_m^{(j-1)} \left(\frac{\boldsymbol{a}_m^H \Sigma^{-1} \boldsymbol{R}_{\boldsymbol{Y}} \Sigma^{-1} \boldsymbol{a}_m}{\boldsymbol{a}_m^H \Sigma^{-1} \boldsymbol{a}_m} \right)}$$

Stepsize $\mu \in [0,1]$ determines γ convergence

Main result: Convergence $\lim_{j\to\infty} \gamma_m^{(j)}$ is guaranteed...

- for any $\mu \in [0,1]$ if $\gamma_m = 0$, i.e., $\forall m \notin \mathcal{M}$
- for some $\mu \in [0, \mu_{\max}]$ if $\gamma_m > 0$, i.e., $\forall m \in \mathcal{M}$, with $\mu_{\max} < 1$

Contents

Introduction

Array Data Model

M-Estimation based on CES-distribution

Sketch of the Proof

Simulation Results

Conclusion

Sketch of the Proof (1/5)

The proof proceeds in two steps

1. The true solution is a fixed point of the iteration
2. The iteration update decreases the error

According to the update rule:

$$\gamma_m^{(j)} = (1-\mu)\gamma_m^{(j-1)} + \mu\gamma_m^{(j-1)} \left(\frac{\boldsymbol{a}_m^H \Sigma^{-1} \boldsymbol{R}_{\boldsymbol{Y}} \Sigma^{-1} \boldsymbol{a}_m}{\boldsymbol{a}_m^H \Sigma^{-1} \boldsymbol{a}_m} \right)$$

and assuming perfect estimate $\boldsymbol{R}_{\boldsymbol{Y}} = \Sigma$ we see that

$$\boxed{\boldsymbol{\gamma}^{(j)} = \boldsymbol{\gamma}^{\mathrm{true}} \quad \text{if} \quad \boldsymbol{\gamma}^{(j-1)} = \boldsymbol{\gamma}^{\mathrm{true}}}$$

where $\boldsymbol{\gamma}^{\mathrm{true}}$ is the true vector of source powers for any μ.

The estimate $\boldsymbol{R}_{\boldsymbol{Y}}$ is unbiased for Σ when $\boldsymbol{y}_\ell$ follow any of distributions in the Table (thanks to b).

For convergence:

the new error must be less than the previous error,

$$|\gamma_m^{\mathrm{true}} - \gamma_m^{(j)}| \quad < \quad |\gamma_m^{\mathrm{true}} - \gamma_m^{(j-1)}|.$$

This is ensured if

$$\left| 1 + \mu\gamma_m^{(j-1)} \left(\frac{\partial G_m(\gamma)}{\partial \gamma_m} \right)_{\gamma=\tilde{\gamma}} \right| < 1,$$

where

$$G_m(\gamma) = \frac{\boldsymbol{a}_m^H \Sigma^{-1} \boldsymbol{R}_{\boldsymbol{Y}} \Sigma^{-1} \boldsymbol{a}_m}{\boldsymbol{a}_m^H \Sigma^{-1} \boldsymbol{a}_m}$$

For inactive indices m, we have $\gamma_m^{\mathrm{true}} = 0$,

and it turns out that any μ with $0 < \mu < 1$ guarantees $\gamma_m^{(j)} \to 0$.

The active set is estimated for any μ

The active set determines DOAs

For active indices m, we have $\gamma_m^{\mathrm{true}} > 0$,

and it turns out that any μ with $0 < \mu < \mu_{\mathsf{max}}$ guarantees $\gamma_m^{(j)} \to \gamma_m^{\mathrm{true}}$.

Source power estimation requires smaller μ than DOA estimation

Contents

Introduction

Array Data Model

M-Estimation based on CES-distribution

Sketch of the Proof

Simulation Results

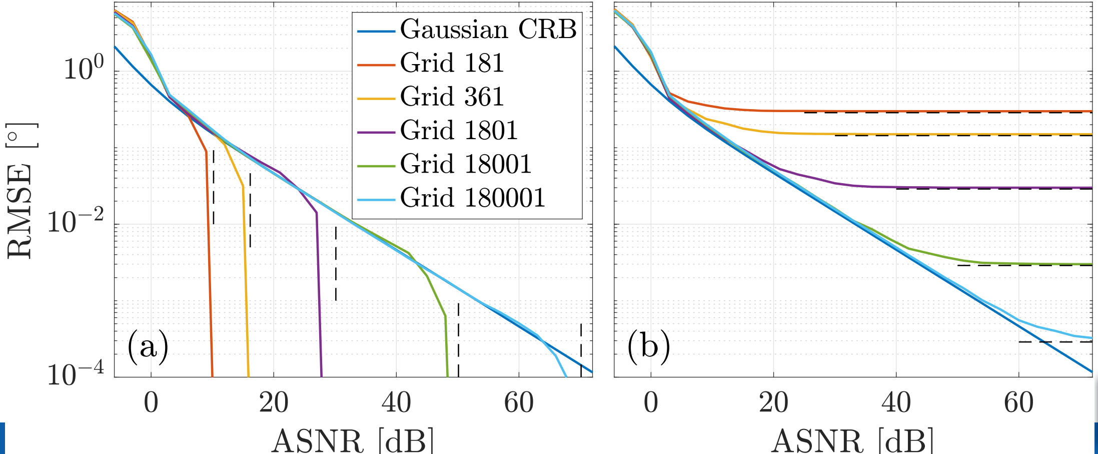
Conclusion

Simulations Results

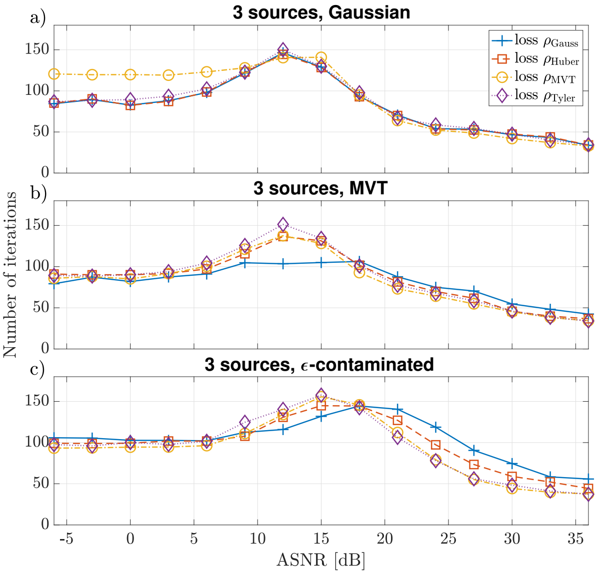
- Simulation with synthetic data for ULA,
  - $N = 20$ sensors,
  - $L = 25$ array data snapshots,
  - $K = 1$ a single active source at DOA $\theta = -45^\circ$
  - Two DOA grid sizes $M = 181$ (low res) and $M = 18001$ (high res).
- We evaluate Root Mean Square Error (RMSE) of DOA estimators vs. Array Signal to Noise Ratio (ASNR).
- Results averaged over $10^6/L = 4 \cdot 10^4$ realizations.
- Several noise models:
  - Complex multivariate Gaussian
  - Complex multivariate Student, $\nu = 2.1$
  - ϵ-contaminated complex multivariate Gaussian, $\epsilon = 0.05$, $\lambda = 10$

Simulation Results (Grid Resolution Effect on RMSE)

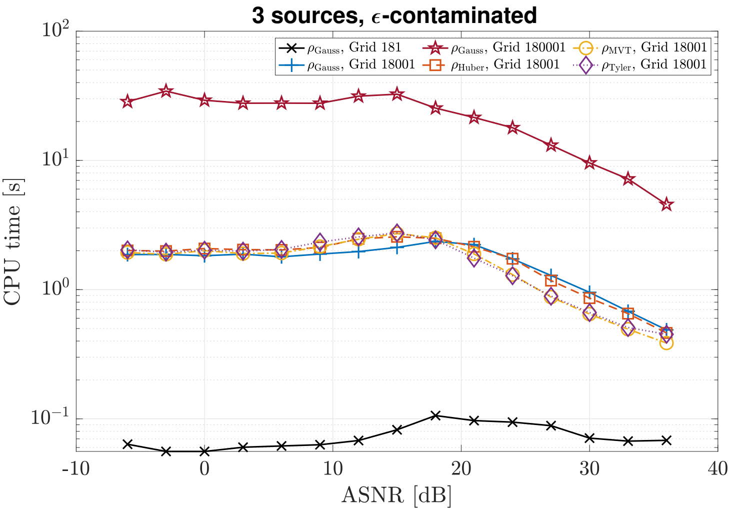
left: fixed DOA -10° on grid, right: uniformly distributed DOA $\sim -10^\circ + U(-\delta/2, \delta/2)$.



Simulation Results (Array Data Distribution Effect on Required Iteration Count)



Simulation Results (Grid Resolution Effect on Required Iteration Count)



Contents

Introduction

Array Data Model

M-Estimation based on CES-distribution

Sketch of the Proof

Simulation Results

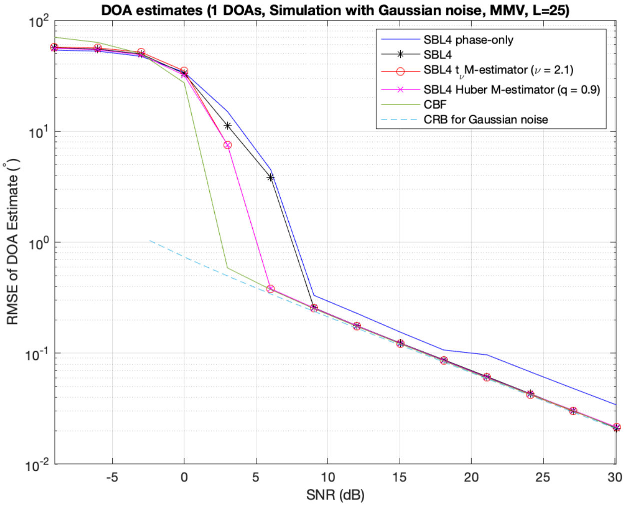
Conclusion

Summary and Conclusion

- Convergence of an M-Estimator for DOAs investigated for CES array data
- General approach based on loss functions
- Investigation of iteration convergence: Analytically and Numerically
-
- M-Estimator DOA performs well for all investigated noise models
- Matlab implementation on GitHub
  https://github.com/NoiseLabUCSD/RobustSBL

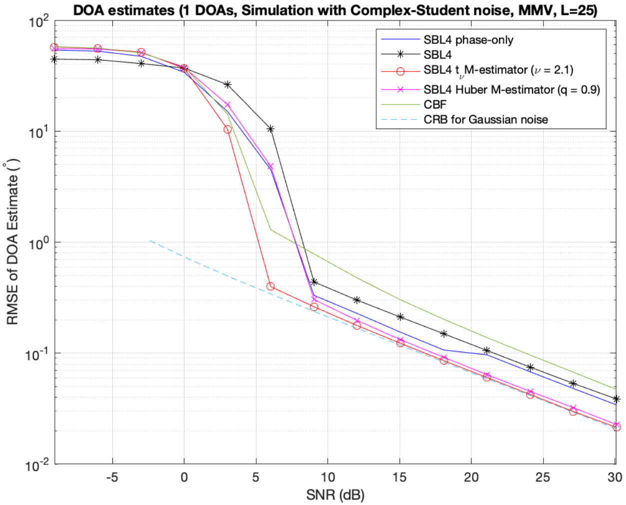
Backup: Simulation Results (Gaussian noise)

$M = 18001$, DOA grid spacing 0.01°



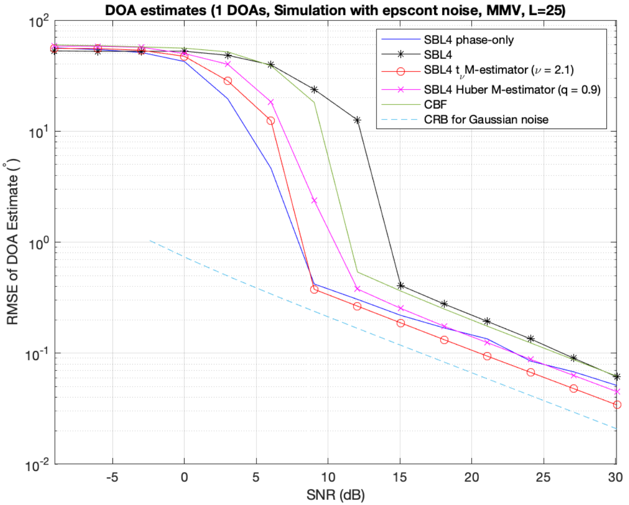
Backup: Simulation Results (Complex Student noise, $\nu = 2.1$)

$M = 18001$, DOA grid spacing 0.01°



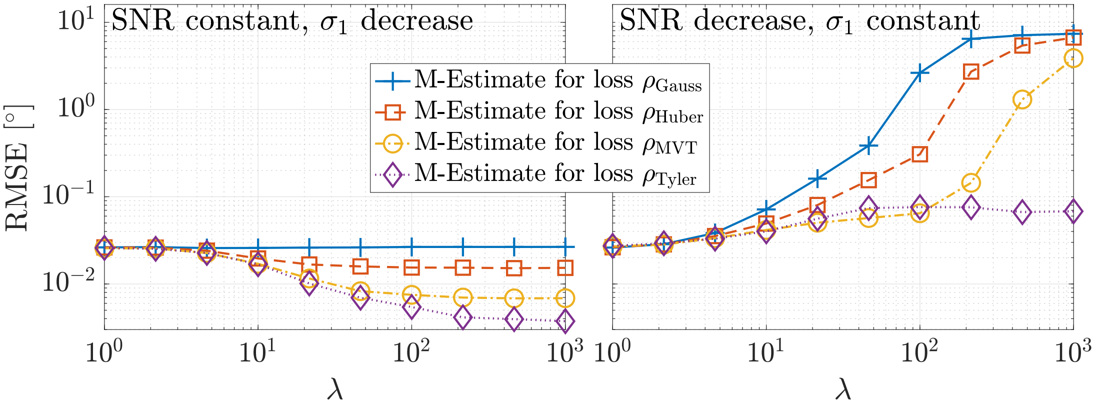
Backup: Simulation Results (ϵ-contaminated noise, $\epsilon = 0.05$, $\lambda = 10$)

$M = 18001$, DOA grid spacing 0.01°



Backup: Simulation Results (Outlier Strength Effect on RMSE)

ϵ-contaminated data $\epsilon = 0.05$ with one source, high SNR



Backup: DOA M-Estimation Algorithm

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

Table: DOA M-Estimation using Sparse Bayesian Learning

- The CRB is in [21, Eq. (8.106)].

$$C_{CR}(\boldsymbol{\theta}) = \frac{\sigma^2}{2L}\left\{\mathrm{Re}\left\{\left[\Gamma_{\mathcal{M}}\left[\left(\boldsymbol{I}_K + \boldsymbol{A}_{\mathcal{M}}^H \boldsymbol{A}_{\mathcal{M}} \frac{\Gamma_{\mathcal{M}}}{\sigma^2}\right)^{-1}\left(\boldsymbol{A}_{\mathcal{M}}^H \boldsymbol{A}_{\mathcal{M}} \frac{\Gamma_{\mathcal{M}}}{\sigma^2}\right)\right]\right] \odot \boldsymbol{H}^T\right\}\right\}^{-1}$$

with

$$\begin{aligned}
\Gamma_{\mathcal{M}} &= \mathrm{diag}(\boldsymbol{\gamma}_{\mathcal{M}}),\\
\boldsymbol{H} &= \boldsymbol{D}^H\left(\boldsymbol{I}_N - \boldsymbol{A}_{\mathcal{M}}\boldsymbol{A}_{\mathcal{M}}^{+}\right)\boldsymbol{D},\\
\boldsymbol{D} &= \left[\left(\frac{\partial \boldsymbol{a}(\theta)}{\partial\theta}\right)_{\theta=\theta_1} \cdots \left(\frac{\partial \boldsymbol{a}(\theta)}{\partial\theta}\right)_{\theta=\theta_K}\right].
\end{aligned}$$

References I

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

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

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

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

References II

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

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

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

References III

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

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

[10] A. G. Jaffer.
Maximum likelihood direction finding of stochastic sources: A separable solution.
In IEEE Int. Conf. on Acoust., Speech, and Sig. Process. (ICASSP-88), volume 5, pages 2893–2896, 1988.

References IV

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

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

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

References V

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

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

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

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

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

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

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

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

Convergence Analysis of Robust and Sparse M-Estimation of DOA

Christoph Mecklenbräuker[1], *Peter Gerstoft*[2], *Esa Ollila*[3], *Yongsung Park*[2]

[1]Inst. of Telecommunications, TU Wien, Austria
[2]NoiseLab, UCSD, San Diego, USA
[3]Dept. of Information and Communications Engineering, Aalto Univ., Finland

ITA 2024, San Diego, Feb. 19, 2024

Technische Universität Wien

UCSD

Aalto University