PHYSICALLY PARAMETERIZED DIFFERENTIABLE MUSIC FOR DOA ESTIMATION WITH UNCALIBRATED ARRAYS

Baptiste CHATELIER[‡,†], José Miguel MATEOS-RAMOS[⋆], Vincent CORLAY[‡], Christian HÄGER[⋆], Matthieu CRUSSIERE[†], Henk WYMEERSCH[⋆], Luc LE MAGOAROU[†]

† Univ Rennes, INSA Rennes, CNRS, IETR-UMR 6164, Rennes, France
‡ Mitsubishi Electric R&D Centre Europe, Rennes, France
⋆ Department of Electrical Engineering, Chalmers University of Technology, Sweden

ICC 2025, Montreal – June 10, 2025

MODEL-BASED MACHINE LEARNING

Typical data processing setting:

- We observe a ***large*** number of ***correlated*** variables, explained by a ***small*** number of ***independent*** factors.

MODEL-BASED MACHINE LEARNING

Typical data processing setting:

- We observe a ***large*** number of ***correlated*** variables, explained by a ***small*** number of ***independent*** factors.

There are two complementary approaches to handle this situation:

MODEL-BASED MACHINE LEARNING

Typical data processing setting:

- We observe a ***large*** number of ***correlated*** variables, explained by a ***small*** number of ***independent*** factors.

There are two complementary approaches to handle this situation:

- **Signal processing**
  - Model based
  - Large bias
  - Low complexity

MODEL-BASED MACHINE LEARNING

Typical data processing setting:

- We observe a ***large*** number of ***correlated*** variables, explained by a ***small*** number of ***independent*** factors.

There are two complementary approaches to handle this situation:

- **Signal processing**
  - Model based
  - Large bias
  - Low complexity
- **Machine learning/Artificial intelligence**
  - Data based
  - Low bias
  - High complexity

MODEL-BASED MACHINE LEARNING

Typical data processing setting:

- We observe a ***large*** number of ***correlated*** variables, explained by a ***small*** number of ***independent*** factors.

There are two complementary approaches to handle this situation:

- **Signal processing**
  - Model based
  - Large bias
  - Low complexity
- **Machine learning/Artificial intelligence**
  - Data based
  - Low bias
  - High complexity

Hybrid approach: Model-based AI

Use models to structure, initialize and train learning methods

MODEL-BASED MACHINE LEARNING

Typical data processing setting:

- We observe a ***large*** number of ***correlated*** variables, explained by a ***small*** number of ***independent*** factors.

There are two complementary approaches to handle this situation:

- **Signal processing**
  - Model based
  - Large bias
  - Low complexity
- **Machine learning/Artificial intelligence**
  - Data based
  - Low bias
  - High complexity

Hybrid approach: Model-based AI
Use models to structure, initialize and train learning methods

- Make models more flexible: reduce bias of signal processing methods

MODEL-BASED MACHINE LEARNING

Typical data processing setting:

- We observe a ***large*** number of ***correlated*** variables, explained by a ***small*** number of ***independent*** factors.

There are two complementary approaches to handle this situation:

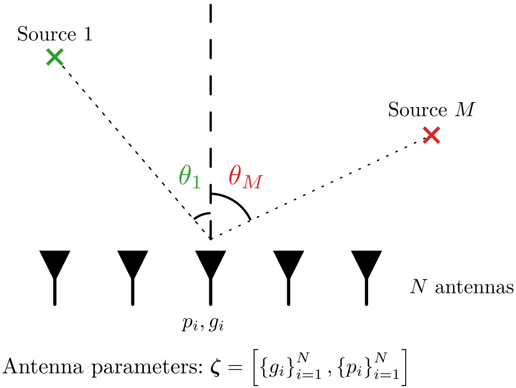
- **Signal processing**
  - Model based
  - Large bias
  - Low complexity
- **Machine learning/Artificial intelligence**
  - Data based
  - Low bias
  - High complexity

Hybrid approach: Model-based AI
Use models to structure, initialize and train learning methods

- Make models more flexible: reduce bias of signal processing methods
- Guide machine learning methods: reduce their complexity

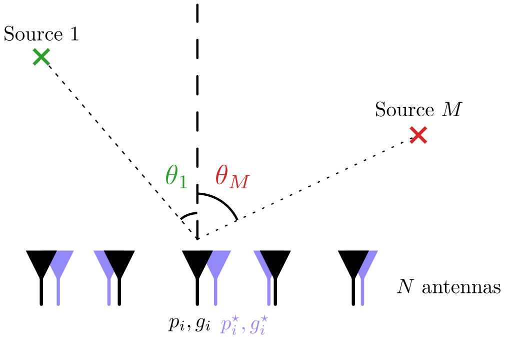
DOA ESTIMATION PROBLEM

- From measurements on N distinct antennas, how to estimate the direction of arrivals $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_M]$ of M non-coherent far-field sources?



DOA ESTIMATION PROBLEM

- From measurements on N distinct antennas, how to estimate the direction of arrivals $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_M]$ of M non-coherent far-field sources, even with an imperfect antenna array?



Antenna parameters: $\boldsymbol{\zeta} = \left[\{g_i\}_{i=1}^N, \{p_i\}_{i=1}^N\right]$

DOA ESTIMATION PROBLEM: MATHEMATICAL FORMULATION

- System model:

$$\mathbf{X} = \mathbf{A}_{\boldsymbol{\zeta}}\left(\boldsymbol{\theta}\right)\mathbf{S} + \mathbf{N} \tag{1}$$

with $\mathbf{X} \in \mathbb{C}^{N\times T}$, $\boldsymbol{\theta} \in [-\pi/2, \pi/2]^{M}$, $\mathbf{A}_{\boldsymbol{\zeta}}\left(\boldsymbol{\theta}\right) \in \mathbb{C}^{N\times M}$, $\mathbf{S} \in \mathbb{C}^{M\times T}$, $\mathbf{N} \in \mathbb{C}^{N\times T}$

DOA ESTIMATION PROBLEM: MATHEMATICAL FORMULATION

- System model:

$$\mathbf{X} = \mathbf{A}_{\boldsymbol{\zeta}}(\boldsymbol{\theta})\,\mathbf{S} + \mathbf{N} \tag{1}$$

with $\mathbf{X} \in \mathbb{C}^{N\times T}, \boldsymbol{\theta} \in [-\pi/2, \pi/2]^M, \mathbf{A}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}) \in \mathbb{C}^{N\times M}, \mathbf{S} \in \mathbb{C}^{M\times T}, \mathbf{N} \in \mathbb{C}^{N\times T}$

  - $N = 16$ antennas (ULA)

DOA ESTIMATION PROBLEM: MATHEMATICAL FORMULATION

- System model:

$$\mathbf{X} = \mathbf{A}_{\boldsymbol{\zeta}}(\boldsymbol{\theta})\,\mathbf{S} + \mathbf{N} \tag{1}$$

with $\mathbf{X} \in \mathbb{C}^{N\times T}$, $\boldsymbol{\theta} \in [-\pi/2, \pi/2]^M$, $\mathbf{A}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}) \in \mathbb{C}^{N\times M}$, $\mathbf{S} \in \mathbb{C}^{M\times T}$, $\mathbf{N} \in \mathbb{C}^{N\times T}$

  - $N = 16$ antennas (ULA)
  - $T = 100$ snapshots

DOA ESTIMATION PROBLEM: MATHEMATICAL FORMULATION

- System model:

$$\mathbf{X} = \mathbf{A}_{\boldsymbol{\zeta}}(\boldsymbol{\theta})\,\mathbf{S} + \mathbf{N} \tag{1}$$

with $\mathbf{X} \in \mathbb{C}^{N\times T}, \boldsymbol{\theta} \in [-\pi/2, \pi/2]^{M}, \mathbf{A}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}) \in \mathbb{C}^{N\times M}, \mathbf{S} \in \mathbb{C}^{M\times T}, \mathbf{N} \in \mathbb{C}^{N\times T}$

  - $N = 16$ antennas (ULA)
  - $T = 100$ snapshots
  - M non-coherent sources

DOA ESTIMATION PROBLEM: MATHEMATICAL FORMULATION

- System model:

$$\mathbf{X} = \mathbf{A}_{\boldsymbol{\zeta}}(\boldsymbol{\theta})\,\mathbf{S} + \mathbf{N} \tag{1}$$

with $\mathbf{X} \in \mathbb{C}^{N\times T}, \boldsymbol{\theta} \in [-\pi/2, \pi/2]^M, \mathbf{A}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}) \in \mathbb{C}^{N\times M}, \mathbf{S} \in \mathbb{C}^{M\times T}, \mathbf{N} \in \mathbb{C}^{N\times T}$

  - $N = 16$ antennas (ULA)
  - $T = 100$ snapshots
  - M non-coherent sources
  - $\boldsymbol{\zeta}$: parameters of the array manifold

DOA ESTIMATION PROBLEM: MATHEMATICAL FORMULATION

- System model:

$$\mathbf{X} = \mathbf{A}_{\zeta}(\boldsymbol{\theta})\,\mathbf{S} + \mathbf{N} \tag{1}$$

with $\mathbf{X} \in \mathbb{C}^{N\times T}, \boldsymbol{\theta} \in [-\pi/2, \pi/2]^M, \mathbf{A}_{\zeta}(\boldsymbol{\theta}) \in \mathbb{C}^{N\times M}, \mathbf{S} \in \mathbb{C}^{M\times T}, \mathbf{N} \in \mathbb{C}^{N\times T}$

  - $N = 16$ antennas (ULA)
  - $T = 100$ snapshots
  - M non-coherent sources
  - ζ: parameters of the array manifold

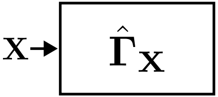
How to estimate θ from $\mathbf{X}$?

MUSIC METHOD

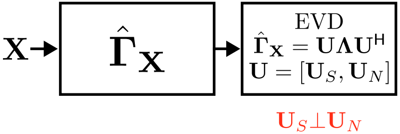
$$\mathbf{X}$$

- Input: measurements

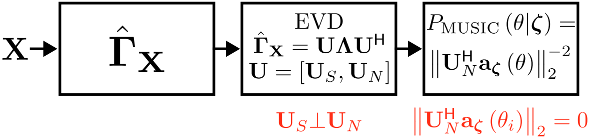
MUSIC METHOD



- Compute the sample covariance matrix from measurements

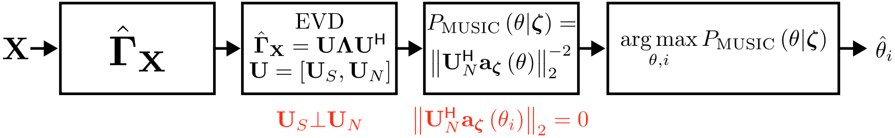


- Apply EVD on the sample covariance matrix

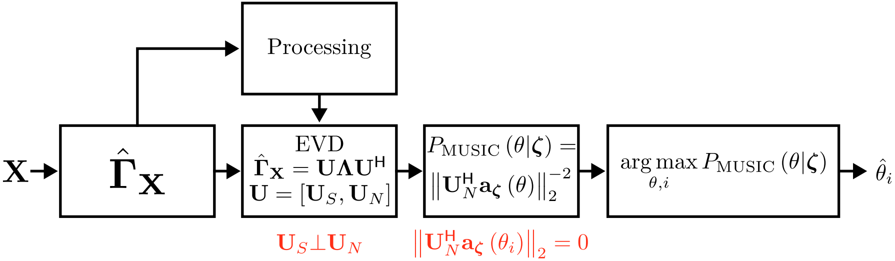


- Compute the MUSIC spectrum

MUSIC METHOD



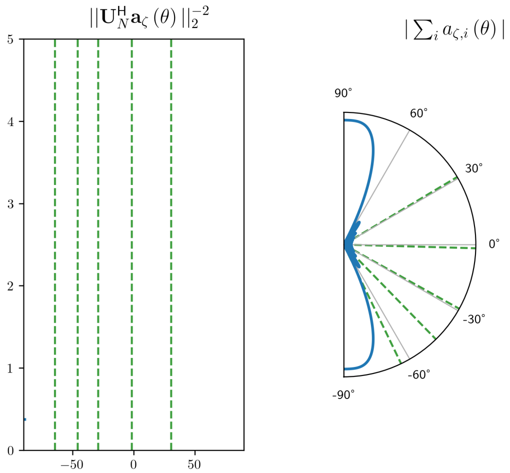
- Find peaks and estimate DoAs



- If the sources are correlated: possibility of finding a surrogate covariance matrix through additional processing[1]

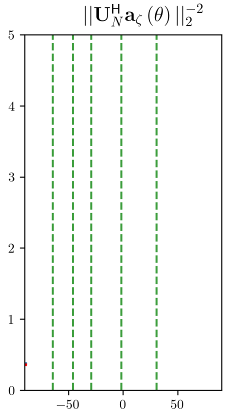
[1] Shmuel et al., “SubspaceNet: Deep Learning-Aided Subspace Methods for DoA Estimation”.

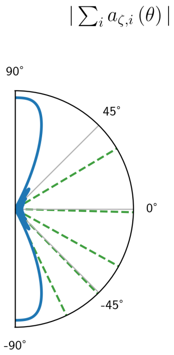
MUSIC METHOD VISUALIZATION



- What happens if ζ is not perfectly known?

MUSIC METHOD VISUALIZATION





Estimation error if ζ is not perfectly known. How to learn ζ?

CONTRIBUTIONS

- **Differentiable MUSIC algorithm to learn HWI through stochastic gradient-descent**

CONTRIBUTIONS

- **Differentiable MUSIC algorithm to learn HWI through stochastic gradient-descent**
- **Problem-specific supervised and unsupervised loss functions**

PROPOSED METHOD: MUSIC NON-DIFFERENTIABLITY

- Main idea: leverage SGD to solve

$$\begin{aligned} \underset{\boldsymbol{\zeta}}{\text{minimize}} \quad & \mathbb{E}_{(\boldsymbol{\theta},\mathbf{X})\sim\mathcal{P}_{(\boldsymbol{\theta},\mathbf{X})}}\left[\mathcal{L}\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\boldsymbol{\zeta}\right)\right)\right], \\ \text{subject to} \quad & \boldsymbol{\zeta}\in\mathbb{C}^N\times\mathbb{R}^N \end{aligned} \tag{P1}$$

PROPOSED METHOD: MUSIC NON-DIFFERENTIABLITY

- Main idea: leverage SGD to solve

$$\begin{aligned} \underset{\boldsymbol{\zeta}}{\text{minimize}} \quad & \mathbb{E}_{(\boldsymbol{\theta},\mathbf{X})\sim\mathcal{P}_{(\boldsymbol{\theta},\mathbf{X})}}\left[\mathcal{L}\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\boldsymbol{\zeta}\right)\right)\right], \\ \text{subject to} \quad & \boldsymbol{\zeta} \in \mathbb{C}^N \times \mathbb{R}^N \end{aligned} \tag{P1}$$

- Requires computing: $\nabla_{\boldsymbol{\zeta}}\mathcal{L}\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\boldsymbol{\zeta}\right)\right)$

PROPOSED METHOD: MUSIC NON-DIFFERENTIABLITY

- Main idea: leverage SGD to solve

$$\begin{aligned} \underset{\boldsymbol{\zeta}}{\text{minimize}} \quad & \mathbb{E}_{(\boldsymbol{\theta},\mathbf{X})\sim\mathcal{P}_{(\boldsymbol{\theta},\mathbf{X})}}\left[\mathcal{L}\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)\right)\right], \\ \text{subject to} \quad & \boldsymbol{\zeta}\in\mathbb{C}^N\times\mathbb{R}^N \end{aligned} \tag{P1}$$

- Requires computing: $\nabla_\zeta\mathcal{L}\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)\right)=\nabla_{\hat{\boldsymbol{\theta}}(\mathbf{X}|\zeta)}\mathcal{L}\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)\right)\nabla_\zeta\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)$

PROPOSED METHOD: MUSIC NON-DIFFERENTIABLITY

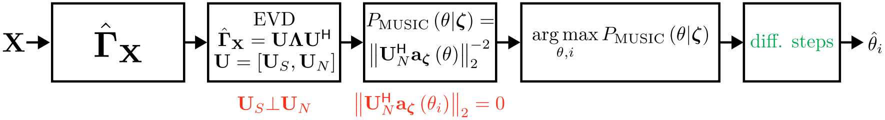
- Main idea: leverage SGD to solve

$$\begin{aligned} \underset{\zeta}{\text{minimize}} \quad & \mathbb{E}_{(\boldsymbol{\theta},\mathbf{X})\sim\mathcal{P}_{(\boldsymbol{\theta},\mathbf{X})}} \left[\mathcal{L}\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)\right)\right], \\ \text{subject to} \quad & \zeta \in \mathbb{C}^N \times \mathbb{R}^N \end{aligned} \tag{P1}$$

- Requires computing: $\nabla_\zeta \mathcal{L}\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)\right) = \nabla_{\hat{\boldsymbol{\theta}}(\mathbf{X}|\zeta)} \mathcal{L}\left(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)\right) \nabla_\zeta \hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)$
  - The $\arg\max$ in MUSIC leads to the non-existence of $\nabla_\zeta \hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)$

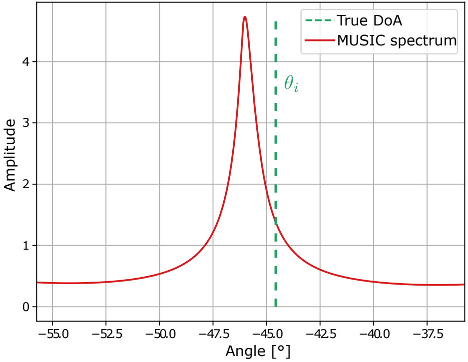
MUSIC is non-differentiable $\rightarrow$ ***diffMUSIC***

PROPOSED METHOD: TOWARDS DIFFMUSIC



- diffMUSIC consists in the addition of differentiable processing steps after the peak-finding method.

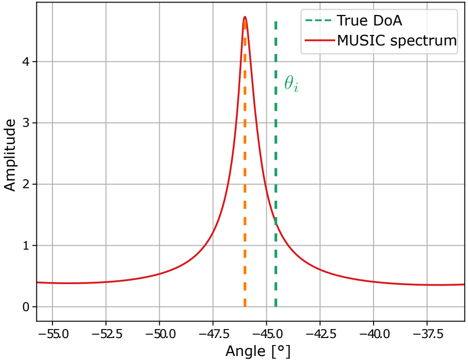
PROPOSED METHOD: DIFFMUSIC DETAILS



- Compute the MUSIC spectrum with current array knowledge ζ: $P_{\text{MUSIC}}\left(\theta|\zeta\right)$

PROPOSED METHOD: DIFFMUSIC DETAILS

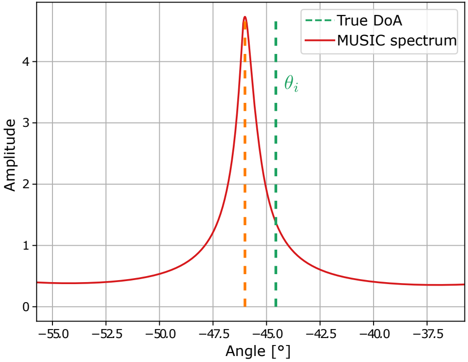
$$\hat{\theta}_i = \underset{\theta,i}{\arg\max}\, P_{\text{MUSIC}}\left(\theta|\zeta\right)$$



- Find peaks in the spectrum

PROPOSED METHOD: DIFFMUSIC DETAILS

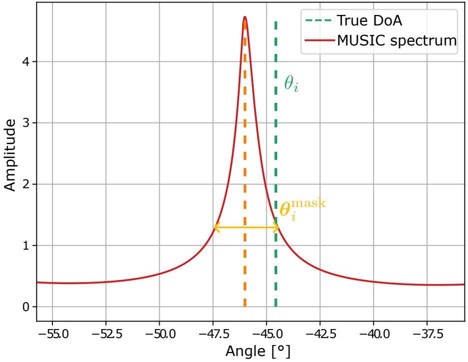
$$\hat{\theta}_i = \underset{\theta,i}{\arg\max}\, P_{\mathrm{MUSIC}}\left(\theta|\zeta\right)$$



- For each peak, estimate the DoA:

PROPOSED METHOD: DIFFMUSIC DETAILS

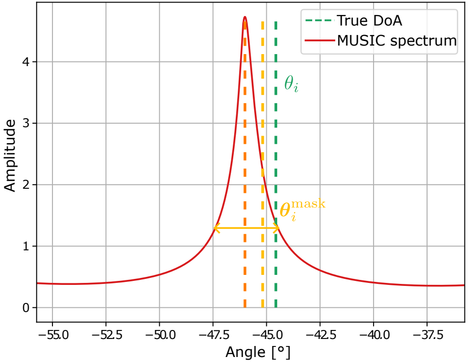
$$\hat{\theta}_i = \arg\max_{\theta, i} P_{\text{MUSIC}}(\theta|\zeta)$$



- For each peak, estimate the DoA:
  - Select neighbor angles through windowing: $\boldsymbol{\theta}_i^{\text{mask}}$

PROPOSED METHOD: DIFFMUSIC DETAILS

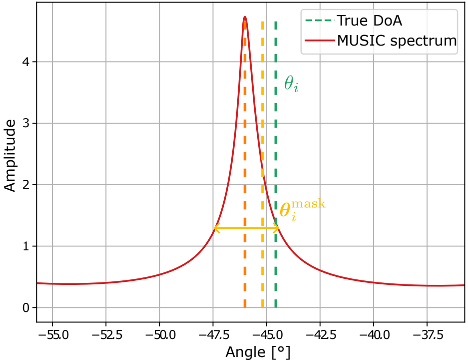
$$\hat{\theta}_i = \left(\boldsymbol{\theta}_i^{\text{mask}}\right)^\top \texttt{softmax}\left(P_{\text{MUSIC}}\left(\boldsymbol{\theta}_i^{\text{mask}}|\zeta\right)\right)$$



- For each peak, estimate the DoA:
  - Select neighbor angles through windowing: $\boldsymbol{\theta}_i^{\text{mask}}$
  - Convex combination: $\nabla_\zeta \hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)$ exists $\rightarrow$ differentiable.

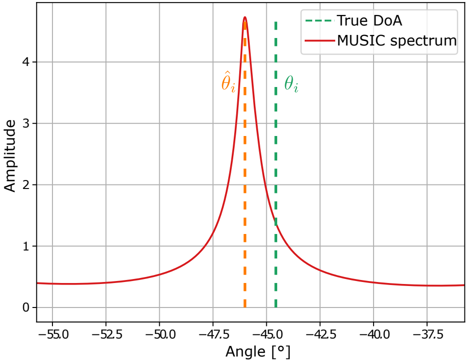
PROPOSED METHOD: DIFFMUSIC DETAILS

$$\hat{\theta}_i = \left(\boldsymbol{\theta}_i^{\text{mask}}\right)^\top \texttt{softmax}\left(P_{\text{MUSIC}}\left(\boldsymbol{\theta}_i^{\text{mask}}|\zeta\right)\right)$$



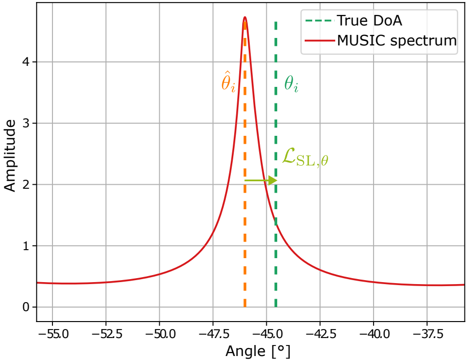
- Update the array parameters: $\zeta \leftarrow \zeta - \mu \nabla_\zeta \mathcal{L}\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\left(\mathbf{X}|\zeta\right)\right)$

PROPOSED METHOD: LOSS FUNCTIONS



- How to design task-adapted loss functions to learn ζ?

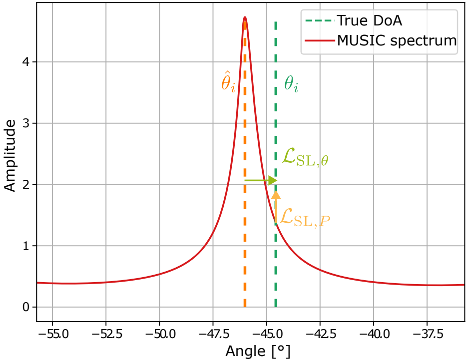
PROPOSED METHOD: LOSS FUNCTIONS



- Minimize the estimation error: RMSPE

$$\mathcal{L}_{\mathrm{SL},\theta} = \frac{1}{|\mathcal{T}|} \sum_{(\boldsymbol{\theta},\mathbf{X})\in\mathcal{T}} \min_{\mathbf{P}\in\mathcal{P}} \sqrt{\frac{1}{M} \left\| \mathrm{mod}_{\pi} \left(\boldsymbol{\theta} - \mathbf{P}\hat{\boldsymbol{\theta}} \left(\mathbf{X} | \zeta \right) \right) \right\|_2^2} \quad (3)$$

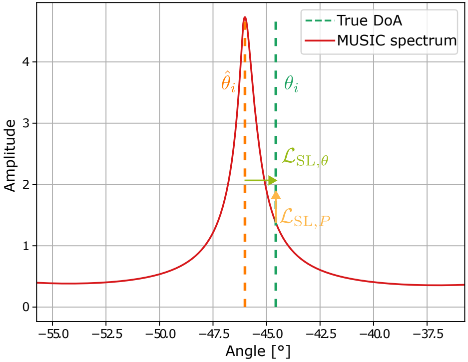
PROPOSED METHOD: LOSS FUNCTIONS



- Maximize spectrum amplitude at true DoA locations:

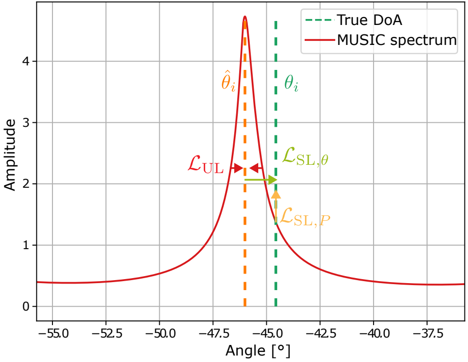
$$\mathcal{L}_{\mathrm{SL},P} = -\frac{1}{|\mathcal{T}|} \sum_{(\boldsymbol{\theta},\mathbf{X})\in\mathcal{T}} \sum_{i} P_{\mathrm{MUSIC}}\left(\theta_i|\boldsymbol{\zeta}\right) \tag{3}$$

PROPOSED METHOD: LOSS FUNCTIONS



Requires a-priori knowledge of the true DoAs!

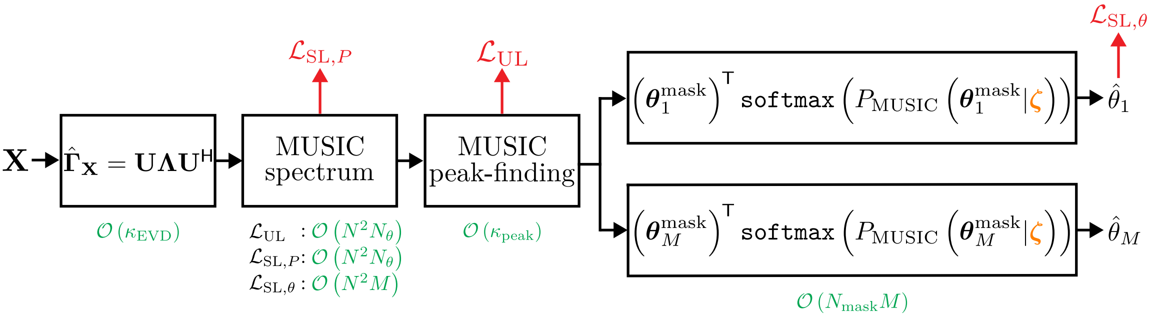
PROPOSED METHOD: LOSS FUNCTIONS



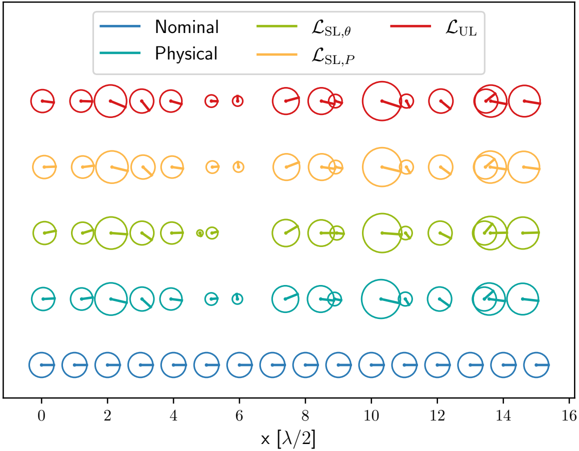
- Unsupervised learning: maximize spectrum sharpness within the chosen angular window (Jain's index based)

$$\mathcal{L}_{\mathrm{UL}} = \frac{1}{|\mathcal{T}|} \sum_{\mathbf{X} \in \mathcal{T}} \sum_{i} \mathfrak{J}\left(P_{\mathrm{MUSIC}}\left(\boldsymbol{\theta}_i^{\mathrm{mask}}\left(\mathbf{X}|\zeta\right)|\zeta\right)\right) \tag{3}$$

PROPOSED METHOD: LOSS FUNCTIONS

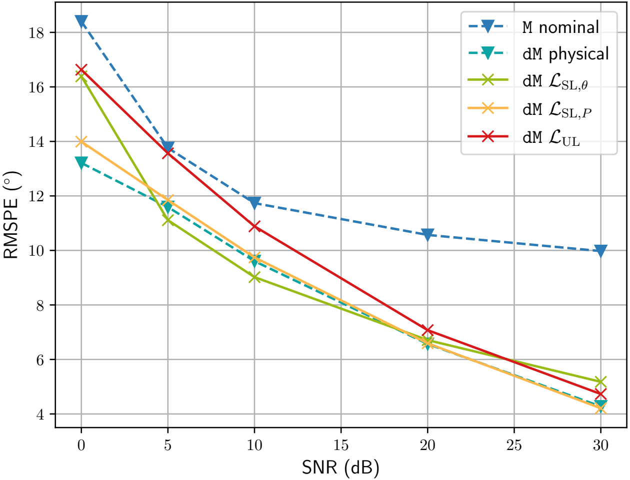


EXPERIMENTAL RESULTS: $N = 16, M = 5$, HIGH HWIS



- The proposed method learns the impairments

EXPERIMENTAL RESULTS: $N = 16$, $M = 5$, HIGH HWIS



- The proposed method performs well under noise

EXPERIMENTAL RESULTS: PERFORMANCE AGAINST BASELINES

		Baselines				$\mathcal{L}_{\mathrm{SL},\theta}$		$\mathcal{L}_{\mathrm{SL},P}$		$\mathcal{L}_{\mathrm{UL}}$	
		M (nom.)	M (phys.)	dM (phys.)	SubspaceNet	M	dM	M	dM	M	dM
RMSPE ($^\circ$)	$M = 1$	2.425	0.014	0.013	0.098	0.019	0.015	0.013	$\mathbf{0.013}$	1.339	1.310
	$M = 5$	9.976	4.358	4.275	16.123	5.371	5.178	4.325	$\mathbf{4.209}$	4.834	4.731

Baseline comparisons

- The proposed method outperforms classical MUSIC and SubspaceNet

CONCLUSION

- Contributions:
  - MUSIC can be modified to be differentiable
  - HWIs can be learned while performing DoA estimation
  - Better results than MUSIC with unknown impairments

CONCLUSION

- Contributions:
  - MUSIC can be modified to be differentiable
  - HWIs can be learned while performing DoA estimation
  - Better results than MUSIC with unknown impairments

- Future work:
  - Extend the method to coherent sources $\rightarrow$ spatial augmentation method
  - Extend the method to near-field $\rightarrow$ new dictionary expression
  - Combine with `SubspaceNet` $\rightarrow$ learn both the surrogate covariance matrix and the array parameters

Thanks!