Subspace Representation Learning for Sparse Linear Arrays to Localize More Sources than Sensors: A Deep Learning Methodology

Kuan-Lin Chen and Bhaskar D. Rao

Department of Electrical and Computer Engineering
University of California, San Diego

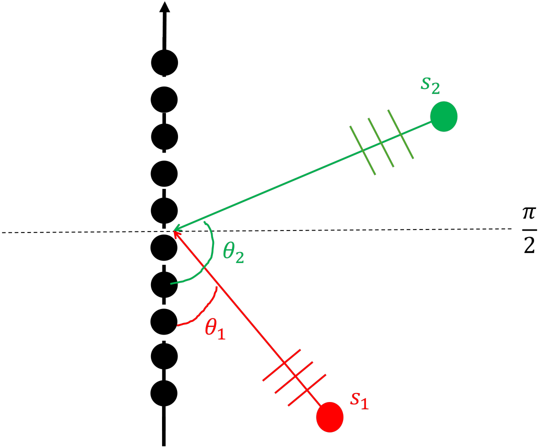
August 2024

Paper is available at https://doi.org/10.1109/TSP.2025.3544170.
Code is available at https://github.com/kjason/SubspaceRepresentationLearning.

Outline

Outline

The direction of arrival (DoA) estimation problem



A uniform linear array (ULA)

Assumptions

- Under the standard assumptions, the snapshot $\mathbf{y}(t) \in \mathbb{C}^M$ at time $t \in [T]$ can be modeled as

$$\mathbf{y}(t) = \sum_{i=1}^{k} s_i(t)\mathbf{a}(\theta_i) + \mathbf{n}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t), \quad \mathbf{n}(t) \sim \mathcal{CN}(\mathbf{0}, \eta \mathbf{I}_M) \quad (1)$$

  where $\mathbf{a}(\theta) : [0, \pi] \to \mathbb{C}^M$ is the array manifold of the M-element uniform linear array (ULA) whose i-th element is given by

$$[\mathbf{a}(\theta)]_i = e^{j2\pi\left(i-1-\frac{(M-1)}{2}\right)\frac{d}{\lambda}\cos\theta}, i \in [M] \quad (2)$$

  and $\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix}\mathbf{a}(\theta_1) & \mathbf{a}(\theta_2) & \cdots & \mathbf{a}(\theta_k)\end{bmatrix}$. The k signals have equal powers.
- Given $\{\mathbf{y}(t)\}_{t=1}^{T}$ and $k \in [M-1]$, how to find $\theta_1, \theta_2, \cdots, \theta_k$?

Background

DoAs $\theta_1, \theta_2, \cdots, \theta_k$ can be found by subspace methods such as MUtiple SIgnal Classification (MUSIC) (Schmidt, 1986) and root-MUSIC (Barabell, 1983; Rao and Hari, 1989).

- Estimate the SCM $\hat{\mathbf{R}}_0$ from $\hat{\mathbf{R}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}(t)\mathbf{y}^{\mathrm{H}}(t)$.
- Find the signal subspace $\mathbf{E}_s$ or noise subspace $\mathbf{E}_n$ via eigenvalue decomposition

$$\hat{\mathbf{R}}_0 = \begin{bmatrix} \mathbf{E}_s & \mathbf{E}_n \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_k & \\ & \mathbf{\Lambda}_{M-k} \end{bmatrix} \begin{bmatrix} \mathbf{E}_s^{\mathrm{H}} \\ \mathbf{E}_n^{\mathrm{H}} \end{bmatrix}. \tag{3}$$

- MUSIC. Find all the peaks of

$$\frac{1}{\mathbf{a}^{\mathrm{H}}(\theta)\mathbf{E}_n\mathbf{E}_n^{\mathrm{H}}\mathbf{a}(\theta)}. \tag{4}$$

- Root-MUSIC. Find the roots of

$$\mathbf{v}^{\mathrm{H}}\mathbf{E}_n\mathbf{E}_n^{\mathrm{H}}\mathbf{v} = 0 \tag{5}$$

  where $\mathbf{v} = \begin{bmatrix} 1 & z^1 & z^2 & \cdots & z^{(M-1)} \end{bmatrix}^{\mathrm{T}}$.
- An M-element ULA can find $M-1$ sources.

Sparse linear arrays (SLAs)

- Let $N \leq M$ and $\mathcal{S} = \{s_1, s_2, \cdots, s_N\} \subset [M]$. Consider minimum redundancy arrays (MRAs) or nested arrays.[a]
- The snapshot received on this physical array is

$$\mathbf{y}_{\mathcal{S}}(t) = \mathbf{\Gamma}\mathbf{y}(t). \tag{6}$$

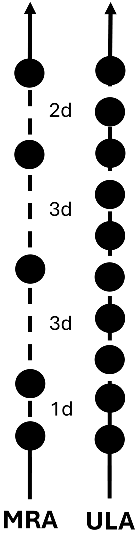
where $\mathbf{\Gamma} \in \mathbb{R}^{N \times M}$ is a row selection matrix given by

$$[\mathbf{\Gamma}]_{nm} = \begin{cases} 1, & \text{if } s_n = m, \\ 0, & \text{otherwise,} \end{cases}, n \in [N], m \in [M]. \tag{7}$$

- The noiseless SCM of the SLA/MRA is

$$\mathbf{R}_{\mathcal{S}} = \mathbf{\Gamma}\mathbf{R}_0\mathbf{\Gamma}^{\mathsf{T}}. \tag{8}$$

- Define $\hat{\mathbf{R}}_{\mathcal{S}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}_{\mathcal{S}}(t)\mathbf{y}_{\mathcal{S}}^{\mathsf{H}}(t)$.



[a] Pal, Piya, and Palghat P. Vaidyanathan. "Nested arrays: A novel approach to array processing with enhanced degrees of freedom." *IEEE Transactions on Signal Processing* 58, no. 8 (2010).

Background

$$\mathbf{R}_0 = \text{Toep}\left(\begin{bmatrix} r(0) \\ r(1) \\ \vdots \\ r(M) \end{bmatrix}\right) = \begin{bmatrix} r(0) & r(1) & \cdots & r(M) \\ r^*(1) & r(0) & \cdots & r(M-1) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(M) & r^*(M-1) & \cdots & r(0) \end{bmatrix}. \quad (9)$$

- $\mathbf{R}_0$ of the M-element ULA can be reconstructed from $\mathbf{R}_{\mathcal{S}}$ of the N-element MRA. For example, for $M = 10$ and $N = 5$,

$$\begin{bmatrix} r(0) & r(2) & r(5) & r(8) & r(9) \\ r^*(2) & r(0) & r(3) & r(6) & r(7) \\ r^*(5) & r^*(3) & r(0) & r(3) & r(4) \\ r^*(8) & r^*(6) & r^*(3) & r(0) & r(1) \\ r^*(9) & r^*(7) & r^*(4) & r^*(1) & r(0) \end{bmatrix}. \quad (10)$$

- Redundancy averaging and direct augmentation.[1]

[1] Pillai, S. Unnikrishna, Yeheskel Bar-Ness, and Fred Haber. "A new approach to array geometry for improved spatial spectrum estimation." *Proceedings of the IEEE* 73, no. 10 (1985).

Outline

1. Estimation of more sources than sensors
2. Covariance matrix reconstruction
   - Optimization-based approaches
   - Deep learning-based approaches
3. Subspace representation learning
   - Subspace representations of different dimensions
   - Geodesic distances
   - Approximation guarantees
4. A new gridless end-to-end approach
5. Numerical results

The maximum likelihood problem

- $\mathbf{R}_0 + \eta \mathbf{I}_M$ is positive semidefinite and possibly Toeplitz.
- $\mathbf{y}_\mathcal{S}(t) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_\mathcal{S} + \eta \mathbf{I}_N)$.
- One can formulate the following constrained optimization problem according to the maximum likelihood principle:

$$\begin{aligned} \min_{\mathbf{v} \in \mathbb{C}^M} \quad & \log\det\left(\mathbf{\Gamma}\,\mathrm{Toep}(\mathbf{v})\mathbf{\Gamma}^\mathsf{T}\right) + \mathrm{tr}\left(\left(\mathbf{\Gamma}\,\mathrm{Toep}(\mathbf{v})\mathbf{\Gamma}^\mathsf{T}\right)^{-1}\hat{\mathbf{R}}_\mathcal{S}\right) \\ \text{subject to} \quad & \mathrm{Toep}(\mathbf{v}) \succeq 0. \end{aligned} \tag{11}$$

Convex relaxation and majorization-minimization:

- SPA (Yang et al., 2014), Wasserstein distance minimization (Wang et al., 2019), StructCovMLE (Pote and Rao, 2023), etc

The sparse and parametric approach (SPA)

Based on the covariance fitting criterion (Stoica et al., 2010)[2], Yang et al. (2014)[3] formulated the SPA which involves the following problem:

$$\begin{aligned} \min_{\mathbf{X}\in\mathbb{H}^N, \mathbf{v}\in\mathbb{C}^M} \quad & \operatorname{tr}(\mathbf{X}) + \operatorname{tr}\left(\hat{\mathbf{R}}_{\mathcal{S}}^{-1}\boldsymbol{\Gamma}\mathrm{Toep}(\mathbf{v})\boldsymbol{\Gamma}^{\mathsf{T}}\right) \\ \text{subject to} \quad & \begin{bmatrix} \mathbf{X} & \hat{\mathbf{R}}_{\mathcal{S}}^{\frac{1}{2}} & \\ \hat{\mathbf{R}}_{\mathcal{S}}^{\frac{1}{2}} & \boldsymbol{\Gamma}\mathrm{Toep}(\mathbf{v})\boldsymbol{\Gamma}^{\mathsf{T}} & \\ & & \mathrm{Toep}(\mathbf{v}) \end{bmatrix} \succeq 0. \end{aligned} \tag{12}$$

[2]Stoica, Petre, Prabhu Babu, and Jian Li. "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data." *IEEE Transactions on Signal Processing* 59, no. 1 (2010).

[3]Yang, Zai, Lihua Xie, and Cishen Zhang. "A discretization-free sparse and parametric approach for linear array signal processing." *IEEE Transactions on Signal Processing* 62, no. 19 (2014).

Majorization-minimization

Because the log det term in (11) is concave, it can be majorized by a supporting hyperplane. Based on the majorization-minimization principle, Pote and Rao (2023)[4] proposed the "StructCovMLE" approach that solves a sequence of SDP problems.

- Let $\mathbf{R}^{(0)} = \mathbf{I}_M$.
- For $i = 0, 1, 2, \cdots$, $\mathbf{R}^{(i+1)} = \mathrm{Toep}(\mathbf{v}^*)$ where $\mathbf{v}^*$ is found by solving

$$\begin{aligned} \min_{\mathbf{v}\in\mathbb{C}^M, \mathbf{X}\in\mathbb{H}^N} \quad & \mathrm{tr}\left(\left(\mathbf{\Gamma}\mathbf{R}^{(i)}\mathbf{\Gamma}^\mathsf{T}\right)^{-1}\mathbf{\Gamma}\mathrm{Toep}(\mathbf{v})\mathbf{\Gamma}^\mathsf{T}\right) + \mathrm{tr}\left(\mathbf{X}\hat{\mathbf{R}}_{\mathcal{S}}\right) \\ \text{subject to} \quad & \begin{bmatrix} \mathbf{X} & \mathbf{I}_N & \\ \mathbf{I}_N & \mathbf{\Gamma}\mathrm{Toep}(\mathbf{v})\mathbf{\Gamma}^\mathsf{T} & \\ & & \mathrm{Toep}(\mathbf{v}) \end{bmatrix} \succeq 0. \end{aligned} \tag{13}$$

- Stop the iteration if the relative change of $\mathbf{R}^{(i)}$ and $\mathbf{R}^{(i+1)}$ is sufficiently small.

[4] Pote, Rohan R., and Bhaskar D. Rao. "Maximum likelihood-based gridless DoA estimation using structured covariance matrix recovery and SBL with grid refinement." *IEEE Transactions on Signal Processing* 71 (2023): 802-815.

Grid-based approaches using sparse Bayesian learning

- Pick $G \in \mathbb{N}$ and let $\mathbf{g}$ be a G-point grid of $[0, \pi]$.
- Define $\boldsymbol{\Sigma}(\boldsymbol{\gamma}) = \mathbf{A}(\mathbf{g})\text{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}}(\mathbf{g}) + \lambda \mathbf{I}_M$ for every $\boldsymbol{\gamma} \in \mathbb{R}_+^G$.
- Under the standard setting, the following maximum likelihood problem can be formulated

$$\min_{\boldsymbol{\gamma} \in \mathbb{R}_+^G} \quad \log\det\left(\boldsymbol{\Gamma}\boldsymbol{\Sigma}(\boldsymbol{\gamma})\boldsymbol{\Gamma}^{\mathsf{T}}\right) + \text{tr}\left(\left(\boldsymbol{\Gamma}\boldsymbol{\Sigma}(\boldsymbol{\gamma})\boldsymbol{\Gamma}^{\mathsf{T}}\right)^{-1}\hat{\mathbf{R}}_{\mathcal{S}}\right). \tag{14}$$

- Expectation-maximization algorithms (Wipf and Rao, 2004), Tipping iterations (Tipping, 2001), etc

DNN-based covariance matrix reconstruction[7]

Let $\mathcal{D} = \left\{ \hat{\mathbf{R}}_{\mathcal{S}}^{(l)}, \mathbf{R}_0^{(l)} \right\}_{l=1}^{L}$ be a dataset. Learn a function

$$f_W : \mathbb{C}^{N \times N} \to \mathbb{C}^{M \times M} \tag{15}$$

such that

$$f_{W^*}\left(\hat{\mathbf{R}}_{\mathcal{S}}\right) f_{W^*}^{\mathrm{H}}\left(\hat{\mathbf{R}}_{\mathcal{S}}\right) \approx \mathbf{R}_0. \tag{16}$$

Solve

$$\min_{W} \quad \frac{1}{L} \sum_{l=1}^{L} d\left(f_W\left(\hat{\mathbf{R}}_{\mathcal{S}}^{(l)}\right) f_W^{\mathrm{H}}\left(\hat{\mathbf{R}}_{\mathcal{S}}^{(l)}\right), \mathbf{R}_0^{(l)} \right). \tag{17}$$

where $d : \mathbb{C}^{M \times M} \times \mathbb{C}^{M \times M} \to [0, \infty)$ is a distance. For example,

$$d_{\mathrm{Fro}}\left(\mathbf{E}, \mathbf{F}\right) = \left\| \mathbf{E} - \mathbf{F} \right\|_F, \tag{18}$$

and

$$d_{\mathrm{Aff}}\left(\mathbf{E}, \mathbf{F}\right) = \left\| \log\left(\mathbf{E}^{-\frac{1}{2}} \mathbf{F} \mathbf{E}^{-\frac{1}{2}} \right) \right\|_F.^{56} \tag{19}$$

[5] Given a matrix **B**, another matrix **A** is said to be a matrix logarithm of B if $e^{\mathbf{A}} = \mathbf{B}$.

[6] $\mathbf{R}_0^{(l)}$ is replaced by $\mathbf{R}_0^{(l)} + \delta \mathbf{I}_M$.

[7] Barthelme, Andreas, and Wolfgang Utschick. "DoA estimation using neural network-based covariance matrix reconstruction." *IEEE Signal Processing Letters* 28 (2021).

Toeplitz prior

Learn a function

$$f_W : \mathbb{C}^{N\times N} \to \mathbb{C}^M \tag{20}$$

such that

$$\text{Toep}\left(f_{W^*}\left(\hat{\mathbf{R}}_{\mathcal{S}}\right)\right) \approx \mathbf{R}_0. \tag{21}$$

The squared loss function

$$d_{\text{squ}}(\mathbf{u}, \mathbf{v}) = \frac{1}{2M}\|\mathbf{u} - \mathbf{v}\|_2^2 \tag{22}$$

can be used to train the DNN models.[8]

[8] Wu, Xiaohuan, Xu Yang, Xiaoyuan Jia, and Feng Tian. "A gridless DOA estimation method based on convolutional neural network with Toeplitz prior." *IEEE Signal Processing Letters* 29 (2022): 1247-1251.

The invariance issue in covariance matrix fitting

- The matrix $\alpha \mathbf{R}_0$ should also be a solution for any $\alpha > 0$ because only the signal or noise subspace is needed for the root-MUSIC algorithm.
- The signal subspace can remain unchanged even though the eigenvalues of $\mathbf{R}_0$ are changed.
- However, the above covariance matrix reconstruction methods do not take this invariance into account.
- In fact, covariance matrix reconstruction is a more difficult problem than reconstructing subspaces.

Question 1

Is it possible for a neural network to learn subspaces of different dimensions?

Outline

Subspace representation learning[9]

- Let $\mathcal{D} = \left\{ \hat{\mathbf{R}}_{\mathcal{S}}^{(l)}, \mathcal{U}^{(l)} \right\}_{l=1}^{L}$ be a dataset. Construct

$$f_W : \mathbb{C}^{N \times N} \times [M-1] \rightarrow \bigcup_{k=1}^{M-1} \mathrm{Gr}(k, M) \tag{23}$$

where $\mathrm{Gr}(k, M)$ is the Grassmann manifold or *Grassmannian* such that

$$f_{W^*}\left(\hat{\mathbf{R}}_{\mathcal{S}}, k\right) \approx \mathcal{U}. \tag{24}$$

- Solve

$$\min_{W} \quad \frac{1}{L} \sum_{l=1}^{L} d_{k=k^{(l)}} \left(f_W \left(\hat{\mathbf{R}}_{\mathcal{S}}^{(l)}, k^{(l)} \right), \mathcal{U}^{(l)} \right) \tag{25}$$

where $d_k : \mathrm{Gr}(k, M) \times \mathrm{Gr}(k, M) \rightarrow [0, \infty)$ is some distance.

[9] Chen, Kuan-Lin, and Bhaskar D. Rao. "Subspace Representation Learning for Sparse Linear Arrays to Localize More Sources than Sensors: A Deep Learning Methodology." *IEEE Transactions on Signal Processing* (2025).

Loss functions

- We propose to construct $d_k : \mathrm{Gr}(k, M) \times \mathrm{Gr}(k, M) \to [0, \infty)$ as a function of the vector of *principal angles* between two given subspaces.

- It is a necessary condition[10] if

$$d_k\left(\mathbf{Q} \cdot \mathcal{U}, \mathbf{Q} \cdot \tilde{\mathcal{U}}\right) = d_k\left(\mathcal{U}, \tilde{\mathcal{U}}\right) \tag{26}$$

for every $\mathcal{U}, \tilde{\mathcal{U}} \in \mathrm{Gr}(k, M)$ and every $\mathbf{Q} \in \mathbb{U}(M)$. The *left action* of $\mathbb{U}(M)$ on $\mathrm{Gr}(k, M)$ in (26) is defined by

$$\mathbf{Q} \cdot \mathcal{U} := \mathrm{span}\left(\mathbf{QB}\right) \tag{27}$$

where the columns of $\mathbf{B} \in \mathbb{C}^{M \times k}$ form a basis of $\mathcal{U}$.

[10] Wong, Yung-Chow. "Differential geometry of Grassmann manifolds." *Proceedings of the National Academy of Sciences* 57, no. 3 (1967): 589-594.

Principal angles

- Let $\mathcal{U}, \tilde{\mathcal{U}} \in \mathrm{Gr}(k, M)$ and $\mathbf{U} \in \mathbb{C}^{M \times k}$ and $\tilde{\mathbf{U}} \in \mathbb{C}^{M \times k}$ be matrices whose columns form unitary bases of $\mathcal{U}$ and $\tilde{\mathcal{U}}$, respectively.

- The principal angles $\boldsymbol{\phi}_k = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_k \end{bmatrix}^{\mathsf{T}}$ between $\mathcal{U}$ and $\tilde{\mathcal{U}}$ can be calculated by[11]

$$\phi_i \left(\mathcal{U}, \tilde{\mathcal{U}} \right) = \cos^{-1} \left(\sigma_i \left(\mathbf{U}^{\mathsf{H}} \tilde{\mathbf{U}} \right) \right) \tag{28}$$

  for $i \in [k]$ where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$ are the singular values of $\mathbf{U}^{\mathsf{H}} \tilde{\mathbf{U}}$.

[11] Björck, Åke, and Gene H. Golub. "Numerical methods for computing angles between linear subspaces." *Mathematics of computation* 27, no. 123 (1973): 579-594.

Distances

$$\boldsymbol{\phi}_k = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_k \end{bmatrix}. \qquad \phi_1 \leq \phi_2 \leq \cdots \leq \phi_k. \tag{29}$$

Table 1: Distances between subspaces

Distance	Function of principal angles
Geodesic (arc length)	$\|\boldsymbol{\phi}_k\|_2$
Fubini-Study	$\cos^{-1}\left(\prod_{i=1}^{k}\cos\phi_i\right)$
Chordal (projection Frobenius norm)	$\left(\sum_{i=1}^{k}\sin^2\phi_i\right)^{\frac{1}{2}}$
Projection 2-norm	$\sin\phi_k$
Chordal Frobenius norm	$2\left(\sum_{i=1}^{k}\sin^2\frac{\phi_i}{2}\right)^{\frac{1}{2}}$
Chordal 2-norm	$2\sin\frac{\phi_k}{2}$

The geodesic distance

- Among them, the most natural choice of d_k is the *geodesic distance*[12]

$$d_k^{\mathrm{Geo}}\left(\mathcal{U},\tilde{\mathcal{U}}\right)=\left(\sum_{i=1}^{k}\phi_i^2\left(\mathcal{U},\tilde{\mathcal{U}}\right)\right)^{\frac{1}{2}}=\left\|\phi\left(\mathcal{U},\tilde{\mathcal{U}}\right)\right\|_2 \tag{30}$$

  which defines the length of the shortest curve between two points on the Grassmannian $\mathrm{Gr}(k, M)$.

- The geodesic distance of any two points on $\mathrm{Gr}(k, M)$ is bounded by

$$\sqrt{k}\frac{\pi}{2}. \tag{31}$$

[12] Wong, Yung-Chow. "Differential geometry of Grassmann manifolds." *Proceedings of the National Academy of Sciences* 57, no. 3 (1967): 589-594.

Is it possible for a neural network to learn subspaces?

Theorem 1 (Chen and Rao (2025))

For every $k \in [M-1]$ *and every* $\epsilon > 0$, *there exists a ReLU network* $f : \mathbb{C}^{N \times N} \to Gr(k, M)$ *such that*

$$\int_{[0,\pi]^k} d_k^{Geo}\left(f\left(\mathbf{R}_{\mathcal{S}}\right), P_{\mathbf{A}(\boldsymbol{\theta})}\right) d\boldsymbol{\theta} < \epsilon. \tag{32}$$

- The signal subspace can be generated by evaluating a continuous piecewise linear function at the SCM.
- Proof sketch: continuity of the orthogonal projector

$$\left\{P \in \mathbb{C}^{M \times M} \mid P^{\mathrm{H}} = P, P^2 = P, \operatorname{rank}(P) = k\right\} \tag{33}$$

and

$$\|P_{\mathcal{U}_1} - P_{\mathcal{U}_2}\|_F \to 0 \quad \text{implies} \quad d_k^{\mathrm{Geo}}(\mathcal{U}_1, \mathcal{U}_2) \to 0. \tag{34}$$

Outline

Bypassing the root-MUSIC algorithm

- The earliest end-to-end approach is probably the work by Papageorgiou et al. (2021).[13] It is a grid-based approach.
- The MCENet proposed by Barthelme and Utschick (2021b) is a gridless end-to-end approach. However, it was designed for subarray sampling, not for more sources than sensors.[14]

Question 2

Is it possible to bypass the root-MUSIC algorithm and directly output the angle estimates?

Question 3

Which one is better? The subspace representation or angle representation?

[13] Papageorgiou, Georgios K., Mathini Sellathurai, and Yonina C. Eldar. "Deep networks for direction-of-arrival estimation in low SNR." *IEEE Transactions on Signal Processing* 69 (2021): 3714-3729.

[14] Barthelme, Andreas, and Wolfgang Utschick. "A machine learning approach to DoA estimation and model order selection for antenna arrays with subarray sampling." *IEEE Transactions on Signal Processing* 69 (2021): 3075-3087.

A new gridless end-to-end approach

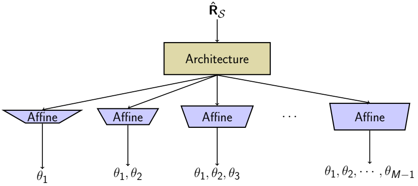


Figure 1: An illustration of the gridless end-to-end model.

We propose to construct a DNN model g_W such that

$$g_W : \mathbb{C}^{N \times N} \times [M-1] \to \mathbb{R}^1 \times \mathbb{R}^2 \times \cdots \mathbb{R}^{M-1}. \tag{35}$$

Solve

$$\min_W \quad \frac{1}{L} \sum_{l=1}^{L} d_{k=k^{(l)}} \left(h_k \circ g_W \left(\hat{\mathbf{R}}_{\mathcal{S}}^{(l)}, k^{(l)} \right), \boldsymbol{\theta}^{(l)} \right) \tag{36}$$

where $d_1, d_2, \cdots, d_{M-1}$ are loss functions of different dimensions that calculate some minimum distances among all permutations. For example, for $k \in [M-1]$,

$$d_k \left(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta} \right) = \frac{1}{k} \min_{\boldsymbol{\Pi} \in \mathcal{P}_k} \left\| \boldsymbol{\Pi} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right\|_2^2. \tag{37}$$

Outline

Experimental setup

- $N = 4,\ M = 7,\ \mathcal{S} = \{1, 2, 5, 7\}$ if not explicitly specified.
- $T = 50.$ $\text{SNR} = 10 \log_{10} \left(\frac{\frac{1}{k} \sum_{i=1}^{k} p_i}{\eta} \right)$.
- $p_1 = p_2 = \cdots = p_k$ if not explicitly specified.
- $k \in [M - 1]$.
- For any $k \in [M - 1]$, the DoAs $\theta_1, \theta_2, \cdots, \theta_k$ are selected at random in the range $\left[\frac{1}{6}\pi, \frac{5}{6}\pi\right]$ with a minimum separation constraint $\min_{i \neq j} |\theta_i - \theta_j| \geq \frac{1}{45}\pi$.

Evaluation and baselines

Evaluation:

- The mean squared error (MSE)

$$\frac{1}{L_{\text{test}}} \sum_{l=1}^{L_{\text{test}}} \frac{1}{k} \min_{\mathbf{\Pi} \in \mathcal{P}_k} \left\| \mathbf{\Pi} \hat{\boldsymbol{\theta}}_l - \boldsymbol{\theta}_l \right\|_2^2 \tag{38}$$

  for a source number $k \in [M-1]$ where $L_{\text{test}} = 10^4$ is the total number of random trials.

Baselines:

- Optimization-based approaches
  - SPA (Yang et al., 2014)
  - Wasserstein distance based approach (WDA) (Wang et al., 2019)
- DNN-based covariance matrix reconstruction (DCR)
  - DCR-G-Fro and DCR-G-Aff (Barthelme and Utschick, 2021a)
  - DCR-T (Wu et al., 2022)

The DNN architecture, dataset, and training procedure

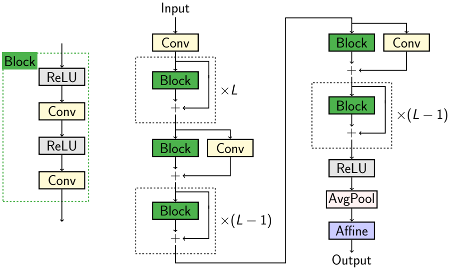


Figure 2: An illustration of a 3-stage L-block ResNet model (He et al., 2016).

- Wide ResNet 16-8 (Zagoruyko and Komodakis, 2016) (11M parameters).
- For each $k \in [M-1]$, there are 2×10^6 and 6×10^5 random data points for training and validation, respectively. $\min_{i \neq j} |\theta_i - \theta_j| \geq \frac{1}{60}\pi$.
- Train 50 epochs with the SGD algorithm and one cycle learning rate scheduler (Smith and Topin, 2019). The batch size is 4096.
- The learning rates for DCR-T, DCR-G-Fro, DCR-G-Aff, and our approach are 0.05, 0.01, 0.005, and 0.1, respectively, found by grid search.

Superior performance over a wide range of SNRs

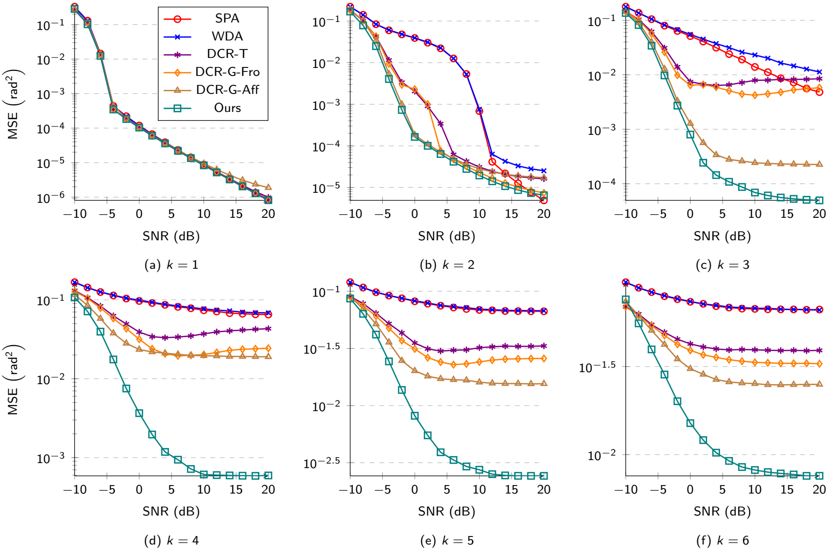


Figure 3: MSE vs. SNR. Only one DNN model is trained for each approach.

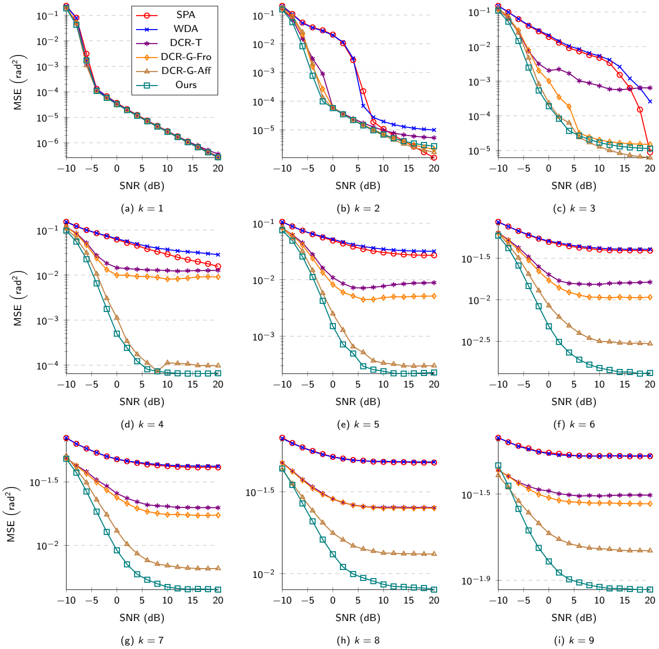
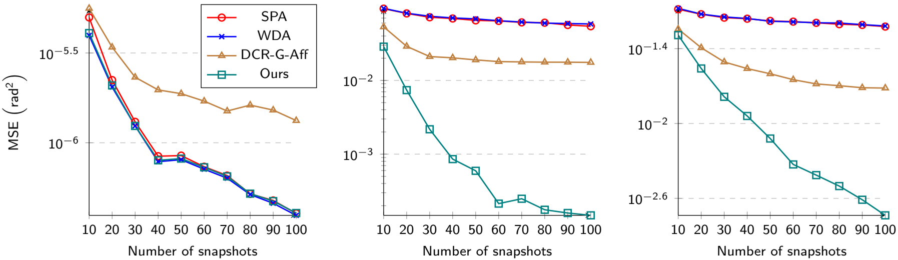


Figure 4: MSE vs. SNR. $N = 5$. $M = 10$.

Performance on unseen numbers of snapshots



(a) $k = 1$ (b) $k = 4$ (c) $k = 6$

Figure 5: MSE vs. number of snapshots. Only one model is trained at 50 snapshots for each approach.

Other distances between subspaces

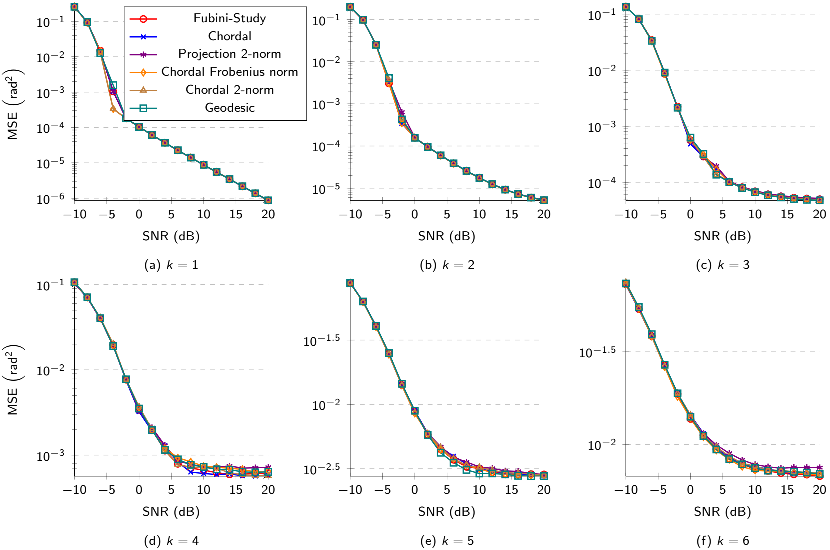


Figure 6: Different distances in Table 1 lead to nearly identical performance.

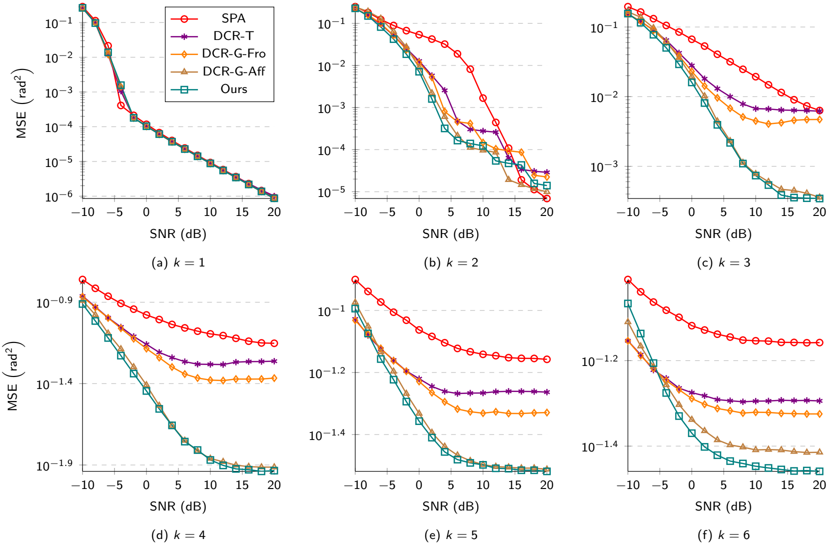


Figure 7: Random source powers with $\frac{\max_i p_i}{\min_j p_j} \leq 10$.

Comparison to the gridless end-to-end approach

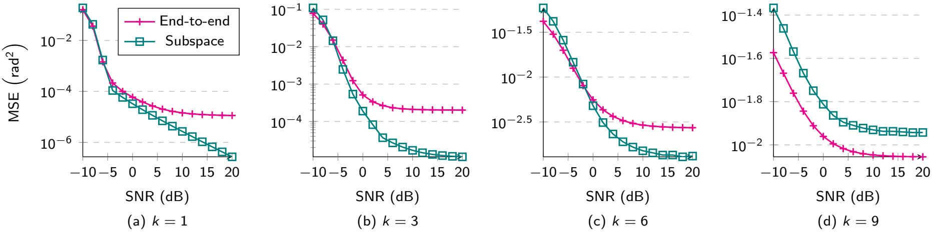


Figure 8: MSE vs. SNR. $N = 5$. $M = 10$.

- For $k \in \{1, 3, 6\}$, subspace representation learning outperforms the gridless end-to-end approach at high SNR regions.
- For $k = 9$, the gridless end-to-end approach is superior.

The imperfect array model[15]

The array manifold with sensor position errors is given by $\mathbf{a}_\rho(\theta) : [0, \pi] \to \mathbb{C}^M$ such that

$$[\mathbf{a}_\rho(\theta)]_i = e^{j2\pi\left(i-1-\frac{(M-1)}{2}+\rho e_i\right)\frac{d}{\lambda}\cos\theta} \tag{39}$$

for $i \in [M]$. The imperfect array manifold $\tilde{\mathbf{a}}_\rho(\theta)$ can be defined as

$$\tilde{\mathbf{a}}_\rho(\theta) = \mathbf{C}_\rho \mathbf{G}_\rho \mathbf{H}_\rho \mathbf{a}_\rho(\theta) \tag{40}$$

where the gain bias is modeled by

$$\mathbf{G}_\rho = \mathbf{I} + \rho \text{diag}\left(g_1, g_2, \cdots, g_M\right), \tag{41}$$

the phase bias is modeled by

$$\mathbf{H}_\rho = \text{diag}\left(e^{j\rho h_1}, e^{j\rho h_2}, \cdots, e^{j\rho h_M}\right), \tag{42}$$

and the intersensor mutual coupling is modeled by

$$\mathbf{C}_\rho = \mathbf{I} + \rho \text{Toep}\left(\begin{bmatrix} 0 & \gamma & \gamma^2 & \cdots & \gamma^{M-1} \end{bmatrix}^\mathsf{T}\right). \tag{43}$$

[15] Liu, Zhang-Meng, Chenwei Zhang, and S. Yu Philip. "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections." *IEEE Transactions on Antennas and Propagation* 66, no. 12 (2018): 7315-7327.

- A larger ρ makes the imperfections more severe.

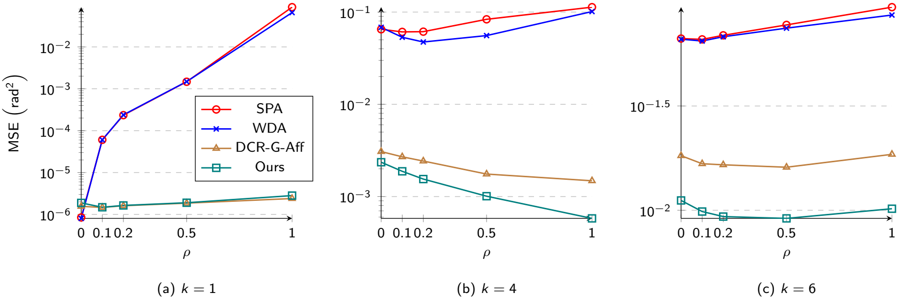


Figure 9: MSE vs. the array imperfection parameter ρ. Only one model is trained for each approach.

Summary

- A new construction to learn subspaces of different dimensions

$$f_W : \mathbb{C}^{N\times N} \times [M-1] \to \bigcup_{k=1}^{M-1} \mathrm{Gr}(k, M). \tag{44}$$

- A DNN model can be trained by distances on $\bigcup_{k=1}^{M-1} \mathrm{Gr}(k, M)$ such as the geodesic distances

$$\left\|\phi_k(\mathcal{U}, \tilde{\mathcal{U}})\right\|_2, \quad k \in [M-1]. \tag{45}$$

- The map between the SCM and the signal subspace can be approximated by a ReLU network.
- A new gridless end-to-end approach is proposed.
- Superior performance compared to SDP-based and DNN-based covariance matrix reconstruction methods.
- The method is geometry/imperfection-agnostic.

References

Barabell, A. (1983). Improving the resolution performance of eigenstructure-based direction-finding algorithms. In *ICASSP'83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 8, pages 336–339. IEEE.

Barthelme, A. and Utschick, W. (2021a). DoA estimation using neural network-based covariance matrix reconstruction. *IEEE Signal Processing Letters*, 28:783–787.

Barthelme, A. and Utschick, W. (2021b). A machine learning approach to DoA estimation and model order selection for antenna arrays with subarray sampling. *IEEE Transactions on Signal Processing*, 69:3075–3087.

Chen, K.-L. and Rao, B. D. (2025). Subspace representation learning for sparse linear arrays to localize more sources than sensors: A deep learning methodology. *IEEE Transactions on Signal Processing*, 73:1293–1308.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 770–778. IEEE.

Papageorgiou, G. K., Sellathurai, M., and Eldar, Y. C. (2021). Deep networks for direction-of-arrival estimation in low SNR. *IEEE Transactions on Signal Processing*, 69:3714–3729.

Pote, R. R. and Rao, B. D. (2023). Maximum likelihood-based gridless DoA estimation using structured covariance matrix recovery and SBL with grid refinement. *IEEE Transactions on Signal Processing*, 71:802–815.

Rao, B. D. and Hari, K. S. (1989). Performance analysis of root-music. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):1939–1949.

Schmidt, R. (1986). Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 34(3):276–280.

Smith, L. N. and Topin, N. (2019). Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, volume 11006, pages 369–386. SPIE.

Stoica, P., Babu, P., and Li, J. (2010). New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data. *IEEE Transactions on Signal Processing*, 59(1):35–47.

Tipping, M. E. (2001). Sparse bayesian learning and the relevance vector machine. *Journal of machine learning research*, 1(Jun):211–244.

Wang, M., Zhang, Z., and Nehorai, A. (2019). Grid-less DOA estimation using sparse linear arrays based on Wasserstein distance. *IEEE Signal Processing Letters*, 26(6):838–842.

Wipf, D. P. and Rao, B. D. (2004). Sparse bayesian learning for basis selection. *IEEE Transactions on Signal processing*, 52(8):2153–2164.

Wu, X., Yang, X., Jia, X., and Tian, F. (2022). A gridless DOA estimation method based on convolutional neural network with Toeplitz prior. *IEEE Signal Processing Letters*, 29:1247–1251.

Yang, Z., Xie, L., and Zhang, C. (2014). A discretization-free sparse and parametric approach for linear array signal processing. *IEEE Transactions on Signal Processing*, 62(19):4959–4973.

Zagoruyko, S. and Komodakis, N. (2016). Wide residual networks. In *British Machine Vision Conference (BMVC)*, pages 87.1–87.12. BMVA Press.