A Comparative Study of Invariance-Aware Loss Functions for Deep Learning-based Gridless Direction-of-Arrival Estimation

Kuan-Lin Chen[†], Bhaskar D. Rao
Department of Electrical and Computer Engineering, University of California, San Diego [†] kuc029@ucsd.edu

UC San Diego JACOBS SCHOOL OF ENGINEERING Electrical and Computer Engineering

2025 ICASSP HYDERABAD, INDIA Celebrating Signal Processing

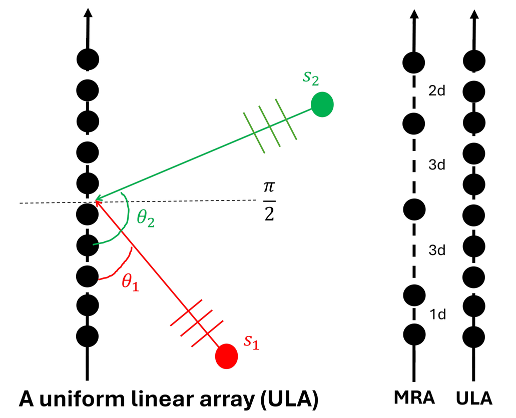
Abstract

- **Covariance matrix reconstruction** has been the most widely used guiding objective in **gridless direction-of-arrival (DoA)** estimation for sparse linear arrays.
- We propose **new loss functions** that are invariant to matrix scaling and study loss functions with **varying degrees of invariance**.
- We provide evidence that designing loss functions with a greater degree of invariance is advantageous.
- Subspace loss from **subspace representation learning** achieves state-of-the-art performance.

1 Gridless DoA estimation

Under the standard assumptions, the snapshot at time $t \in [T]$ is

$$\mathbf{y}(t) = \sum_{i=1}^{k} s_i(t)\mathbf{a}(\theta_i) + \mathbf{n}(t) \quad (1)$$



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

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

Let $n \le m$ and $\mathcal{S} = \{e_1, e_2, \cdots, e_n\} \subset [m]$. Consider a sparse linear array (SLA) such as a minimum redundancy array (MRA) or a nested array. The snapshot received on this physical array is $\mathbf{y}_\mathcal{S}(t) = \mathbf{\Gamma}\mathbf{y}(t)$. where $\mathbf{\Gamma} \in \mathbb{R}^{n\times m}$ is a row selection matrix

$$[\mathbf{\Gamma}]_{ij} = \begin{cases} 1, & \text{if } e_i = j, \\ 0, & \text{otherwise,} \end{cases}, i \in [n], j \in [m]. \quad (3)$$

Let $\mathbf{R}_0$ be the SCM of the ULA. The noiseless SCM of the SLA is $\mathbf{R}_\mathcal{S} = \mathbf{\Gamma}\mathbf{R}_0\mathbf{\Gamma}^\mathsf{T}$. Define $\hat{\mathbf{R}}_\mathcal{S} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_\mathcal{S}(t)\mathbf{y}_\mathcal{S}^\mathsf{H}(t)$.

Question 1. *Find $\theta_1, \theta_2, \cdots, \theta_k$ given $\hat{\mathbf{R}}_\mathcal{S}$ and k.*

- Direct augmentation (DA) [1]:

$$\min_{\mathbf{v}\in\mathbb{C}^m} \left\|\mathbf{\Gamma}\mathrm{Toep}(\mathbf{v})\mathbf{\Gamma}^\mathsf{T} - \hat{\mathbf{R}}_\mathcal{S}\right\|_F. \quad (4)$$

- Sparse and parametric approach (SPA) [2]:

$$\begin{aligned} \min_{\mathbf{X}\in\mathbb{H}^n, \mathbf{v}\in\mathbb{C}^m} \quad & \mathrm{tr}(\mathbf{X}) + \mathrm{tr}\left(\hat{\mathbf{R}}_\mathcal{S}^{-1}\mathbf{\Gamma}\mathrm{Toep}(\mathbf{v})\mathbf{\Gamma}^\mathsf{T}\right) \\ \text{subject to} \quad & \begin{bmatrix} \mathbf{X} & \hat{\mathbf{R}}_\mathcal{S}^{\frac{1}{2}} \\ \hat{\mathbf{R}}_\mathcal{S}^{\frac{1}{2}} & \mathbf{\Gamma}\mathrm{Toep}(\mathbf{v})\mathbf{\Gamma}^\mathsf{T} \\ & \mathrm{Toep}(\mathbf{v}) \end{bmatrix} \succeq 0. \end{aligned} \quad (5)$$

2 DNN-based covariance matrix reconstruction

Training a covariance matrix reconstruction model can be formulated as minimizing the empirical risk

$$\min_W \frac{1}{L}\sum_{l=1}^{L} d\left(g \circ f_W\left(\hat{\mathbf{R}}_\mathcal{S}^{(l)}\right), h\left(\mathbf{R}^{(l)}\right)\right). \quad (6)$$

- $f_W : \mathbb{C}^{n\times n} \to \mathbb{C}^{m\times m}$ is a DNN model with parameters W.
- $\{\hat{\mathbf{R}}_\mathcal{S}^{(l)}, \mathbf{R}^{(l)}\}_{l=1}^{L}$ is a dataset of sample covariance matrices at the SLA and noiseless covariance matrices at the ULA.
- h is a function that extracts the learning target.
- g is a transformation that ensures some properties of a valid covariance matrix. For example, picking the function $g(\mathbf{E}) = \mathbf{E}\mathbf{E}^\mathsf{H} + \delta\mathbf{I}$ for some $\delta \ge 0$ enforces the predicted matrix being always positive semidefinite (or positive definite).
- $d : \mathbb{C}^{m\times m} \times \mathbb{C}^{m\times m} \to [0, \infty)$ is a loss function of choice.

Question 2. *How to design d in (6)?*

2.1 Frobenius norm

$$d_{\mathrm{Fro}}\left(\hat{\mathbf{R}}, \mathbf{R}\right) = \left\|\hat{\mathbf{R}} - \mathbf{R}\right\|_F \quad \text{(Cov).} \quad (7)$$

Remark 1. *$\alpha\mathbf{R}$ for any $\alpha \in \mathbb{R}\setminus\{0\}$ leads to identical signal and noise subspaces.*

Remark 2. *$d_{Fro}(\alpha\mathbf{R}, \mathbf{R}) \to \infty$ as $|\alpha| \to \infty$ for any $\mathbf{R} \succ 0$.*

2.2 Scale-invariant loss functions

To avoid such a penalization and allow a larger solution space, we propose the following *scale-invariant reconstruction loss*:

$$d_{\mathrm{SI}}\left(\hat{\mathbf{R}}, \mathbf{R}\right) = -\log\left(\frac{\|\alpha^*\mathbf{R}\|_F}{\epsilon + \left\|\alpha^*\mathbf{R} - \hat{\mathbf{R}}\right\|_F}\right) \quad \text{(SI-Cov)} \quad (8)$$

where $\epsilon \ge 0$ is a constant and

$$\alpha^* = \arg\min_{\alpha\in\mathbb{R}} \left\|\alpha\mathbf{R} - \hat{\mathbf{R}}\right\|_F. \quad (9)$$

Remark 3. *d_{SI} is invariant to scaling of the matrices in the following sense: For every $\gamma \neq 0$,*

$$d_{SI}(\gamma\mathbf{R}, \mathbf{R}) \to -\infty \quad \text{as} \quad \epsilon \to 0. \quad (10)$$

Remark 4. *Approximately, the property in Remark 3 allows d_{SI} to expand the solution space from a point to a line in $\mathbb{C}^{m\times m}$.*

Remark 5. *If $\epsilon > 0$, in general we have $d_{SI}(\gamma\mathbf{R}_1, \mathbf{R}_1) \neq d_{SI}(\gamma\mathbf{R}_2, \mathbf{R}_2)$ for $\mathbf{R}_1 \neq \mathbf{R}_2$.*

Denoting $\mathbf{E}_s$ as a matrix whose columns are signal eigenvectors of $\mathbf{R}$, the scale-invariant reconstruction loss can be applied to the signal subspace matrix $h(\mathbf{R}) = \mathbf{E}_s\mathbf{E}_s^\mathsf{H}$ as follows

$$-\log\left(\frac{\|\alpha^*\mathbf{E}_s\mathbf{E}_s^\mathsf{H}\|_F}{\epsilon + \left\|\alpha^*\mathbf{E}_s\mathbf{E}_s^\mathsf{H} - g\circ f_W\left(\hat{\mathbf{R}}_\mathcal{S}\right)\right\|_F}\right) \quad (11)$$

where

$$\alpha^* = \arg\min_{\alpha\in\mathbb{R}} \left\|\alpha\mathbf{E}_s\mathbf{E}_s^\mathsf{H} - g\circ f_W\left(\hat{\mathbf{R}}_\mathcal{S}\right)\right\|_F \quad \text{(SI-Sig).} \quad (12)$$

The same formulation can be applied to the noise subspace, where $h(\mathbf{R}) = \mathbf{E}_n\mathbf{E}_n^\mathsf{H}$ and $\mathbf{E}_n$ denotes the noise subspace.

2.3 The affine invariant distance

The affine invariant distance [3]

$$d_{\mathrm{Aff}}\left(\hat{\mathbf{R}}, \mathbf{R}\right) = \left\|\log\left(\mathbf{R}^{-\frac{1}{2}}\hat{\mathbf{R}}\mathbf{R}^{-\frac{1}{2}}\right)\right\|_F \quad \text{(Cov-Aff)} \quad (13)$$

measures the length of the shortest curve between two positive definite matrices.

Proposition 1. *For every m-by-m Hermitian matrix such that $\mathbf{R} \succ 0$ and for every $\alpha > 0$,*

$$d_{\mathrm{Aff}}(\alpha\mathbf{R}, \mathbf{R}) = \sqrt{m}|\log\alpha|. \quad (14)$$

Remark 6. *A logarithmic growth in terms of scaling, much slower than the linear rate of the Frobenius norm.*

Remark 7. *d_{Aff} is not scale-invariant, its increased distance is invariant to the underlying matrix and only depends on the scaling factor, unlike the Frobenius norm, which depends on the matrix.*

Remark 8. *For $\mathbf{R}_1 \neq \mathbf{R}_2$, we have $d_{Aff}(\alpha\mathbf{R}_1, \mathbf{R}_1) = d_{Aff}(\alpha\mathbf{R}_2, \mathbf{R}_2)$, ensuring the same penality for perfect fittings.*

3 Subspace representation learning

Loss functions with the greatest degrees of invariance are perhaps the ones proposed in the *subspace representation learning* methodology [4] that avoids reconstructing covariance matrices.

Construct

$$f_W : \mathbb{C}^{n\times n} \times [m-1] \to \bigcup_{k=1}^{m-1} \mathrm{Gr}(k, m) \quad (15)$$

where $\mathrm{Gr}(k, m)$ is the *Grassmannian* such that

$$f_{W^*}\left(\hat{\mathbf{R}}_\mathcal{S}, k\right) \approx \mathcal{U} \quad (16)$$

where $\mathcal{U}$ is the corresponding signal or noise subspace. Let $\mathcal{D} = \left\{\hat{\mathbf{R}}_\mathcal{S}^{(l)}, \mathcal{U}^{(l)}\right\}_{l=1}^{L}$ be a dataset. Solve

$$\min_W \frac{1}{L}\sum_{l=1}^{L} d_{k=k^{(l)}}\left(f_W\left(\hat{\mathbf{R}}_\mathcal{S}^{(l)}, k^{(l)}\right), \mathcal{U}^{(l)}\right) \quad (17)$$

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

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 $\phi_1, \phi_2, \cdots, \phi_k$ between $\mathcal{U}$ and $\tilde{\mathcal{U}}$ are given by

$$\phi_i\left(\mathcal{U}, \tilde{\mathcal{U}}\right) = \cos^{-1}\left(\sigma_i\left(\mathbf{U}^\mathsf{H}\tilde{\mathbf{U}}\right)\right) \quad (18)$$

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

$$d_{\mathrm{Gr}-k}(\mathcal{U}_1, \mathcal{U}_2) = \sqrt{\sum_{i=1}^{k}\phi_i^2(\mathcal{U}_1, \mathcal{U}_2)} \quad (19)$$

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

4 Numerical results

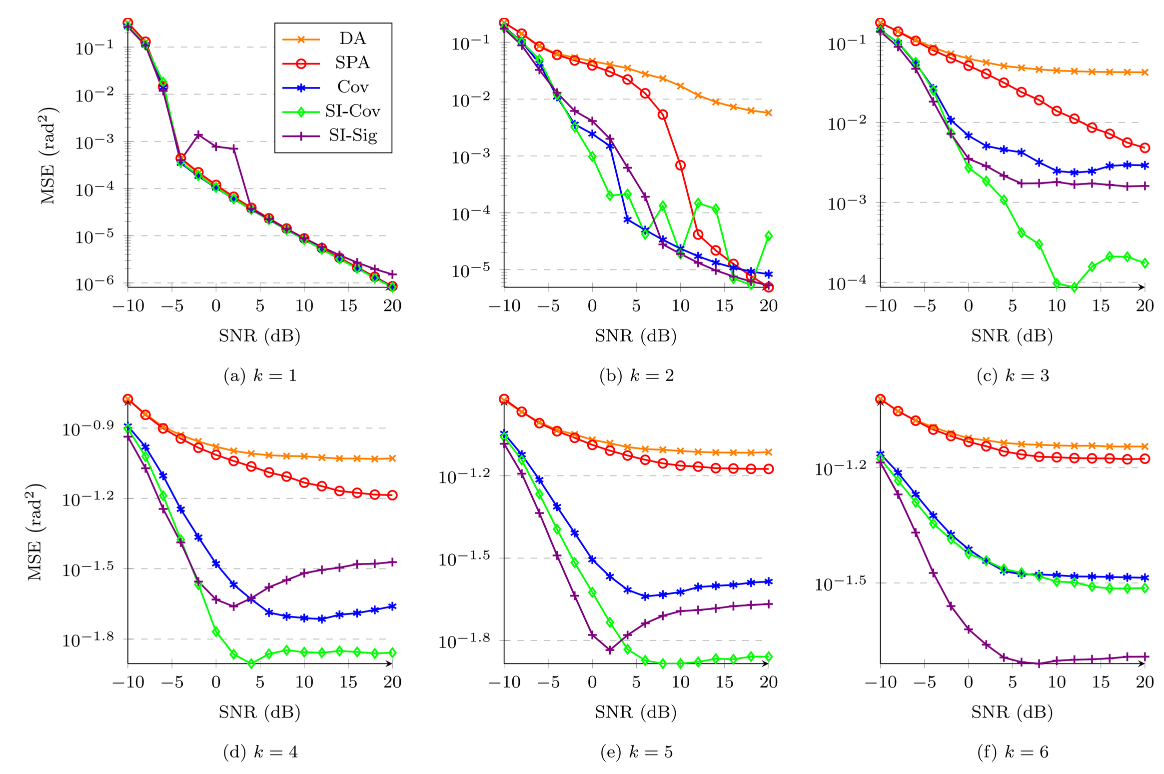


Figure 1: MSE vs. SNR. The proposed scale-invariant covariance matrix reconstruction approach (SI-Cov) outperforms DA, SPA, and Cov when $k > 2$.

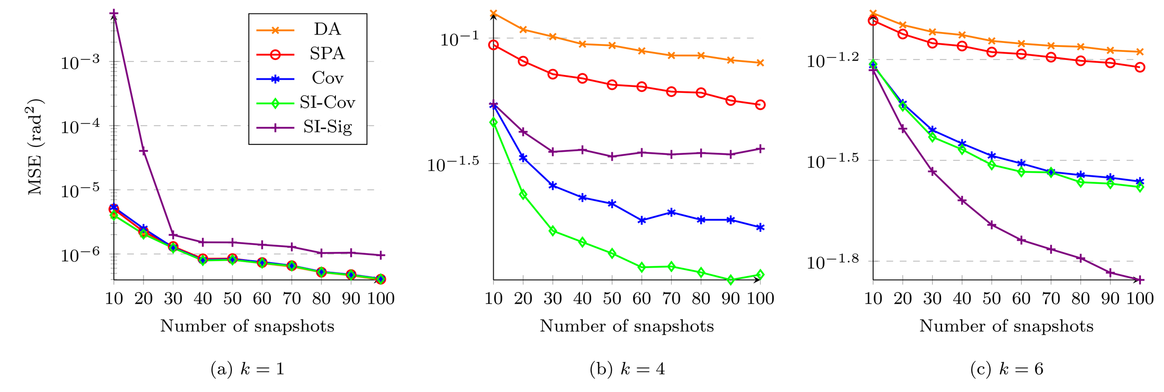


Figure 2: MSE vs. number of snapshots. The proposed SI-Cov outperforms Cov, showing the advantage of using the scale-invariant strategy.

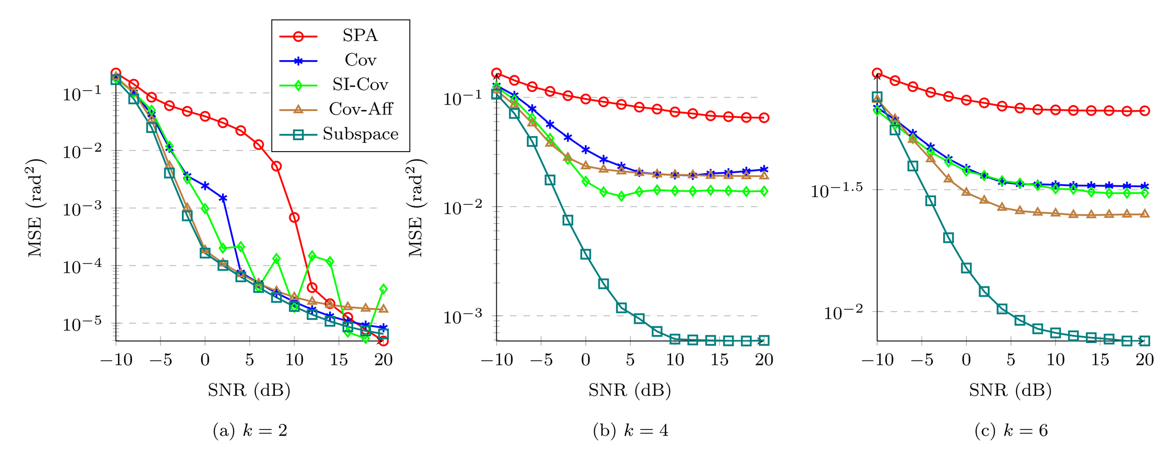


Figure 3: Subspace learning outperforms all the other methods.

Acknowledgements

This work was supported in part by IEEE Signal Processing Society Scholarship, and in part by NSF under Grant CCF-2225617 and Grant CCF-2124929.

References

[1] S. U. Pillai, Y. Bar-Ness, and F. Haber, "A new approach to array geometry for improved spatial spectrum estimation," *Proceedings of the IEEE*, vol. 73, no. 10, pp. 1522–1524, 1985.

[2] Z. Yang, L. Xie, and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4959–4973, 2014.

[3] A. Barthelme and W. Utschick, "DoA estimation using neural network-based covariance matrix reconstruction," *IEEE Signal Processing Letters*, vol. 28, pp. 783–787, 2021.

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