DEEP AUGMENTED MUSIC ALGORITHM FOR DATA-DRIVEN DOA ESTIMATION

Julian P. Merkofer, Guy Revach, Nir Shlezinger, and Ruud J. G. van Sloun

ABSTRACT

Direction of arrival (DoA) estimation is a crucial task in sensor array signal processing, giving rise to various successful model-based (MB) algorithms as well as recently developed data-driven (DD) methods. This paper introduces a new hybrid MB/DD DoA estimation architecture, based on the classical multiple signal classification (MUSIC) algorithm. Our approach augments crucial aspects of the original MUSIC structure with specifically designed neural architectures, allowing it to overcome certain limitations of the purely MB method, such as its inability to successfully localize coherent sources. The deep augmented MUSIC algorithm is shown to outperform its unaltered version with a superior resolution.

***Index Terms*—** DoA estimation, MUSIC, deep learning

1. INTRODUCTION

Source separation and localization are crucial tasks in sensor array processing. In particular, direction of arrival (DoA) estimation of multiple, possibly coherent signal sources plays a key role in a wide range of applications including radar, communications, image analysis, and speech enhancement [1].

A multitude of different DoA estimation algorithms have been proposed, and the problem is still an active area of research. The conventional beamformer is an early approach to DoA estimation, and an extension of classical Fourier-based spectral analysis [2]. Various improvements, such as the minimum variance distortionless response beamformer [3], can alleviate some of its important limitations, such as its inability to successfully estimate multiple closely spaced sources. A leading scheme employed in many DoA applications is the popular multiple signal classification (MUSIC) algorithm [4], which can provide asymptotically unbiased estimates of the number of incident wavefronts present, their approximate frequencies, and their DoA. However, these classical approaches are based on knowledge of the underlying statistical model, and have important limitations. Among the limitations of the model-based (MB) approach is the failure to resolve closely spaced signals with an insufficient number of samples or low signal-to-noise ratio (SNR) scenarios, as well as the inability to consistently estimate the DoA of coherent signals.

The recent success of data-driven (DD) deep learning across a wide range of disciplines gave rise to neural network (NN)-aided DoA estimators. The works [5, 6] implemented model-agnostic DoA estimation using dense and convolutional NNs, respectively. While such black-box NNs enable handling array imperfections due to their model-agnostic nature, they involve highly parameterized models that may be computationally intensive and lack the interpretability of MB methods. Alternatively, NNs were used to robustify the MB MUSIC as a form of a hybrid MB/DD system [7]. Specifically, the work [8] proposed to estimate the discretized MUSIC spectrum from the covariance matrix of the measurements through the utilization of multiple convolutional NNs. While this method is more robust to model inaccuracies compared with the original MUSIC algorithm, it experiences the same drawbacks as its MB counterpart because it utilizes its spatial spectrum as a label for training. Another DD approach proposed in [9] considered systems with subarray sampling and uses NNs to obtain a single estimated covariance matrix from incoherent subarray measurements. This NN-aided estimate is utilized for DoA recovery via the subspace-based MUSIC algorithm. The method addresses the fundamental dependency of MUSIC on the estimated covariance matrix, thereby robustifying the MUSIC algorithm, yet it does not fully exploit the NNs' ability to improve MUSIC, as the NN is trained using the true covariance matrix as a label.

In this work we propose a hybrid MB/DD implementation of MUSIC, exploiting the structure of the classic MUSIC algorithm, while augmenting it with NNs to learn to enhance its performance. Our design builds upon the insight that the sensitivity to model mismatch and inability to handle coherent sources of MUSIC lie in its estimation of the noise and signal subspaces through an eigenvalue decomposition of the empirical covariance matrix. Accordingly, the proposed deep augmentation improves this crucial step by obtaining a pseudo covariance matrix through a NN from the measurements, which is learned along with a NN that acts as a peak finder; i.e., it maps the learned MUSIC spectrum into a set of DoAs while facilitating the learning capabilities of the first NN. As opposed to previous MB/DD variants of MUSIC, the proposed architecture overcomes the fundamental limitations of MUSIC, enabling accurate detection of multiple, possibly coherent sources, while sharing the increased resolution and rapid inference of the MB method. Our experimental study demonstrates the ability of the proposed augmented MUSIC to notably improve upon MB DoA estimators as well as other previously proposed DD DoA estimators. In particular, we

J. P. Merkofer and G. Revach are with the D-ITET, ETH Zurich, Switzerland, (email: julianme@ee.ethz.ch, grevach@ethz.ch). N. Shlezinger is with the School of ECE, Ben-Gurion University of the Negev, Beer Sheva, Israel (e-mail: nirshl@bgu.ac.il). R. J. G. van Sloun is with the EE Dpt., Eindhoven University of Technology, and with Phillips Research, Eindhoven, The Netherlands (e-mail: r.j.g.v.sloun@tue.nl).

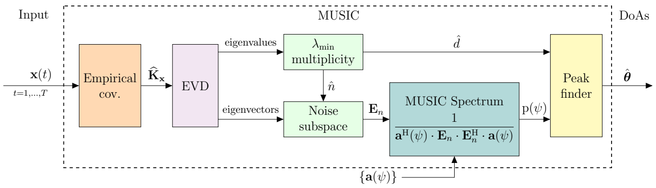

Fig. 1: Block diagram of the MB MUSIC algorithm.

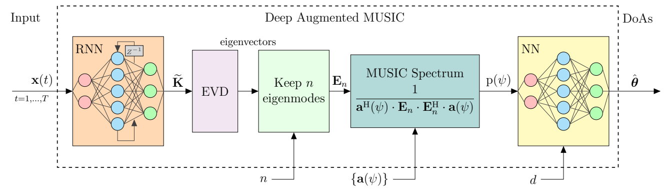

Fig. 2: Block diagram of the augmented MUSIC algorithm.

show that the augmented MUSIC algorithm outperforms its unaltered version in localization accuracy and resolution.

2. SYSTEM MODEL

2.1. Problem Formulation

DoA estimation deals with recovering the incidence angles of signals impinging on a receiving array with m elements, given measurements taken over T time instances obtained as

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta}) \cdot \mathbf{s}(t) + \mathbf{w}(t) \in \mathbb{C}^m, \quad t \in \{1, \ldots, T\}. \quad (1)$$

In the observations model (1), $\mathbf{s}$ represents d narrow-band signals originating from angles $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_d]$, $\mathbf{w} \in \mathbb{C}^m$ is noise, and the matrix $\mathbf{A}(\boldsymbol{\theta})$ comprises *steering* vectors $\mathbf{A}(\boldsymbol{\theta}) = \left[\mathbf{a}(\theta_1) \ \ldots \ \mathbf{a}(\theta_d)\right] \in \mathbb{C}^{m \times d}$. Here, $\mathbf{a}(\theta_i) \in \mathbb{C}^m$ is the steering vector; i.e., the response of each of the m array elements to the incident signal from angle θ_i. For a uniform linear array (ULA) with half-wavelength spacing located in the far-field regions of the sources, the elements of $\mathbf{a}(\theta_i)$ are given by $[\mathbf{a}(\theta_i)]_k = e^{-j\pi k \sin\theta_i}$, where $k \in \{0, \ldots, m-1\}$.

Our goal is to recover the DoAs vector $\boldsymbol{\theta}$ from an observations matrix $\mathbf{X}$ whose columns are T snapshots of the measured waveforms at the m elements $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(T)]$. We assume that we have knowledge of the number of sources d, as well as access to data corresponding to the underlying setup. This data is a set of L pairs, $\{(\mathbf{X}_\ell, \boldsymbol{\theta}_\ell)\}_{\ell=1}^{L}$, each comprising observations and the corresponding DoAs from which they originated. Our design builds upon the MUSIC algorithm formulated in the following section, which estimates the DoAs under additional model assumptions.

2.2. MUSIC Algorithm

The basic idea behind the MUSIC algorithm is to identify the noise and signal subspaces via eigenvalue decomposition (EVD) of the covariance matrix of the received data. These two orthogonal subspaces can then be used to form a pseudo spectrum, whose largest peaks occur at the estimated DoA. To identify the subspaces from the EVD, additional model assumptions are imposed. The sources are assumed to be stationary at all times and non-coherent, such that the covariance matrix $\mathbf{K}_\mathbf{s}$ of $\mathbf{s}$ is diagonal. The noise $\mathbf{w}$ is independent of the source signals and is white, such that its covariance can be written as $\lambda \cdot \mathbf{I}$ for some $\lambda > 0$.

Under the above assumptions, the covariance of the input signal $\mathbf{x}$ is given by $\mathbf{K}_\mathbf{x} = \mathbf{A}(\boldsymbol{\theta}) \cdot \mathbf{K}_\mathbf{s} \cdot \mathbf{A}^{\mathrm{H}}(\boldsymbol{\theta}) + \lambda \cdot \mathbf{I}$. By restricting the number of incident wavefronts d to be less than the number of array elements m, it holds that λ is the minimal eigenvalue of $\mathbf{K}_\mathbf{x}$, denoted λ_{min}, and that its multiplicity is $m - d$. The MB MUSIC algorithm, whose structure is visualized in Fig. 1, exploits this property to estimate the DoAs.

In particular, the MB MUSIC uses the input $\mathbf{X}$ to obtain an empirical estimate $\mathbf{K}_\mathbf{x}$ via $\hat{\mathbf{K}}_\mathbf{x} = \frac{1}{T} \cdot \mathbf{X} \cdot \mathbf{X}^{\mathrm{H}}$. By taking the EVD of $\hat{\mathbf{K}}_\mathbf{x}$, the multiplicity of the eigenvalue λ_{min}, denoted $\hat{n}$, and its corresponding eigenvectors matrix $\mathbf{E}_n \in \mathbb{C}^{m \times \hat{n}}$ are recovered. The former is used to identify the number of incident wavefronts d via $\hat{d} = m - \hat{n}$. The matrix $\mathbf{E}_n$ represents the noise subspace, and thus the squared Euclidean distance from any vector $\mathbf{y} \in \mathbb{C}^m$ to the signal subspace is $\mathbf{y}^{\mathrm{H}} \cdot \mathbf{E}_n \cdot \mathbf{E}_n^{\mathrm{H}} \cdot \mathbf{y}$. Taking the inverse squared distance of $\mathbf{a}(\psi)$ for each ψ gives the spatial spectrum

$$\mathrm{p}_{\mathrm{MUSIC}}(\psi) = \frac{1}{\mathbf{a}^{\mathrm{H}}(\psi) \cdot \mathbf{E}_n \cdot \mathbf{E}_n^{\mathrm{H}} \cdot \mathbf{a}(\psi)}. \quad (2)$$

The DoAs are then estimated by finding the d dominant peaks of (2). A known drawback of the classic MUSIC algorithm is its inability to accurately estimate the DoA angles of coherent signals, as highly correlated signals cause zero entries within the covariance matrix and thereby can become indistinguishable from noise [10]. In such cases, the covariance EVD leads to an imprecise division into noise and signal subspace, even when d is known, thus degrading performance.

3. DEEP AUGMENTED MUSIC

3.1. Architecture

The proposed deep augmented MUSIC algorithm preserves the structure of the MB MUSIC while replacing certain critical aspects with NNs. Our neural augmentation aims to improve the crucial step of estimating the noise and signal subspaces from the empirical covariance, and the translation of the MUSIC spectrum (2) into DoAs via peak finding. By doing so, the proposed augmented MUSIC is not constrained by the additional model assumptions imposed in the derivation of MUSIC in Subsection 2.2, and can, for example, learn to successfully localize coherent signals. This section discusses the overall architecture, while the details of the NNs used in our experimental study are in Section 4.

The augmentation architecture, outlined in Fig. 2, implements the estimation of the noise and signal subspaces in a learned fashion by correlating the measurements of the individual array elements with each other and their different time instances utilizing a recurrent NN (RNN). Since the observations take complex values, as a pre-processing step, the real and imaginary parts of $\mathbf{X}$ are stacked to form a real valued input. The output $\tilde{\mathbf{K}}$, referred to as a pseudo covariance matrix, acts similarly to the covariance matrix of MB MUSIC. However, as its computation is learned from data, it is not constrained to be the empirical covariance. The EVD of $\tilde{\mathbf{K}}$ enables the augmented MUSIC algorithm to categorize the noise and signal subspaces in the same way as the original MUSIC.

Next, the neural augmented MUSIC attains the DoAs from the spatial spectrum $\mathrm{p}(\cdot)$ using an additional NN. The benefits compared to using MB peak finders are two-fold. First, learning the translation of the pseudo-spectrum into DoAs from data enables achieving improved resolution compared to conventional peak finding. Furthermore, peak finding is generally non-differentiable; thus, replacing it with a NN facilitates training the augmented MUSIC end-to-end. The resulting architecture enables the application of gradient-based optimization, by propagating through the NNs as well as the EVD operation, as done in [11]. Doing so allows to jointly tune the noise subspace recovery along with the translation of the MUSIC spectrum into DoAs by comparing its estimated DoAs with the true DoAs.

3.2. Training Procedure

The augmented MUSIC is trained as a multiple regression problem in a supervised fashion. In particular, the system is trained to predict the DoA angles $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \ldots, \hat{\theta}_d]$ by comparing them to the true DoA angles $\boldsymbol{\theta}$ using the root mean squared periodic error (RMSPE) [12]. This measure alters the conventional root mean squared error (RMSE) to accommodate for the periodicity of the DoA angles by computing as the RMSE of the element-wise modulus π error vector. Further, to identify how to match the estimated DoAs $\hat{\boldsymbol{\theta}}$ to the ground truth $\boldsymbol{\theta}$, we compute the loss with respect to the permutation, which minimizes the RMSPE. Consequently, the loss measure used for training the deep augmented MUSIC is

$$\mathrm{RMSPE}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \min_{\mathbf{P} \in \mathcal{P}_d} \left(\frac{1}{d} \left\| \mathrm{mod}_\pi(\boldsymbol{\theta} - \mathbf{P}\hat{\boldsymbol{\theta}}) \right\|^2 \right)^{\frac{1}{2}}, \quad (3)$$

where $\mathcal{P}_d$ is the set of all $d \times d$ permutation matrices.

3.3. Discussion

The proposed augmented MUSIC is customized to overcome certain limitations of MB MUSIC, and particularly those that arise from identifying the noise subspace from the empirical covariance, as well as those associated with the translation of the MUSIC spectrum into DoAs. This is achieved while still maintaining some of the benefits provided by the MB structure. For instance, placing the RNN at the input of the system enables the augmented MUSIC algorithm to operate with any given number of snapshots; i.e., it is invariant of the value of T. This follows because the RNN processes each $\mathbf{x}(t)$ sequentially, thereby allowing real-time tracking of sources.

Furthermore, robustness towards coherent signals is achieved by omitting the calculation of the sample covariance matrix. The integrated EVD facilitates the categorization of the noise and signal subspaces, and can be used to estimate the number of sources d by determining the multiplicity of the smallest eigenvalue as in the original MUSIC algorithm. To enable training end-to-end from the errors in (3), an NN-based peak finder is used, which has a fixed number of outputs. Nonetheless, one can extend the augmented MUSIC to be operable with any number of sources $d < m$; for example, by using gating and/or attention mechanisms to select which outputs are active. We leave this extension of the augmented MUSIC to an unknown number of sources for future work.

Parameter	Value	Parameter	Value
Array elements m	8	Optimizer	Adam [13]
Snapshots T	200	Learning rate	10^{-3}
SNR	10 dB	Weight init	Glorot [14]
Data size L	$9 \cdot 10^4$	Batch size	16

Table 1: Simulation parameters.

Finally, we note that our augmentation approach depends on the array geometry, as the steering vectors $\mathbf{a}(\cdot)$ are used for computing the MUSIC spectrum. Nonetheless, as we numerically demonstrate in Section 4, deep augmented MUSIC learns to overcome mismatches in the array geometry from the data without any alterations or renewed training.

4. EXPERIMENTAL STUDY

4.1. Experimental Setup

We numerically evaluate[1] the proposed deep augmented MUSIC algorithm in comparison to the original MUSIC algorithm as well as the DD DeepMUSIC model [8]. The considered system consists of a ULA with $m = 8$ half wavelength spaced elements. Unless stated otherwise, we use the parameters summarized in Table 1. The training and testing data are generated from the system model (1), where for all $t \in \{1, \ldots, T\}$, the complex-valued $\mathbf{s}_\ell(t)$ and $\mathbf{w}_\ell(t)$ are randomly drawn from the Gaussian distribution $\mathcal{N}(0, 1)$, normalized to meet the fixed SNR. The DoA angles θ_i are drawn uniformly from $(-\frac{\pi}{2}, \frac{\pi}{2})$ in an i.i.d. fashion. In all coherent cases, the sources have identical amplitudes and phases.

The deep augmented MUSIC algorithm illustrated in Fig. 2 is implemented as follows. The input $\mathbf{X}$ consists of the stacked real and imaginary part of the measurements sampled for $T = 200$ time instances and is passed to a gated recurrent unit (GRU). The GRU concurrently processes a single time instance $\mathbf{x}(t)$. After processing all T snapshots, the GRU output is passed to a dense layer, which maps it into the two-dimensional pseudo covariance $\tilde{\mathbf{K}}$, which is transformed back to complex values before being decomposed by an EVD. The spectrum is computed over a uniform grid of 360 angles in $(-\frac{\pi}{2}, \frac{\pi}{2})$ using the estimated noise eigenvectors. This yields the discretized spectrum, which is processed by a multi-layer

[1]The source code used in our numerical study can be found online at `https://github.com/DA-MUSIC/ICASSP22`.

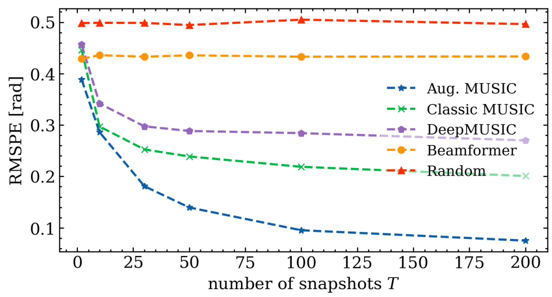

Fig. 3: DoA estimation of $d = 5$ non-coherent signals.

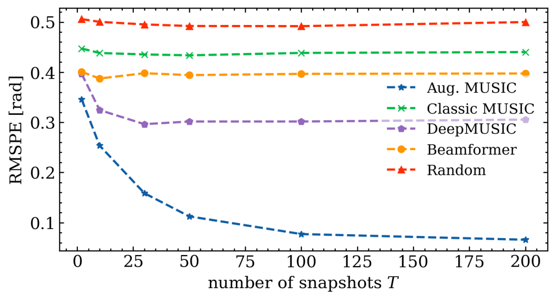

Fig. 4: DoA estimation of $d = 5$ coherent signals.

perceptron (MLP) consisting of three dense layers that extracts the estimated DoA angles $\hat{\boldsymbol{\theta}}$. The number of neurons per layer are chosen in correspondence with the number of sensor elements m and are $2 \cdot m$ for the single layer of the GRU and all three layers of the MLP.

The DD estimators, i.e., our proposed deep augmented MUSIC and DeepMUSIC of [8], are trained for each scenario separately. The classic MUSIC algorithm and the MB beamformer are implemented according to their standard formulations. All presented algorithms have knowledge of d. We implemented DeepMUSIC while incorporating minor alterations that were necessary to accommodate for the difference in the setup compared to that simulated in [8], followed by tuning individual hyperparameters to assure optimal training of the convolutional NNs. The DeepMUSIC architecture is highly parameterized, consisting of approximately 20M trainable parameters, in comparison to the compact deep augmented MUSIC that uses merely 10k parameters.

4.2. Results

The figures show that the deep augmented MUSIC notably outperforms both the MB benchmarks and the DD DeepMUSIC. These gains are observed not only for coherent sources, where the MB MUSIC is known to be inaccurate, but also for non-coherent sources. This indicates the ability of the hybrid MB/DD design to simultaneously improve the accuracy of the MUSIC algorithm as well as allow it to be applied for coherent sources. The accuracy in localizing $d = 5$ sources versus the number of snapshots T available is depicted in Figs. 3-4 for non-coherent and coherent signals, respectively. The deep augmented MUSIC, however, is only trained for $T = 200$, yet through utilization of the GRU, it manages to operate reliably in the low snapshot domain as well. The random algorithm corresponds to the performance when choosing DoA angles at random. Figs. 3-4 exemplify the severe degradation of the MB MUSIC when localizing coherent signals, and highlights the robustness of the deep augmented MUSIC algorithm for such scenarios.

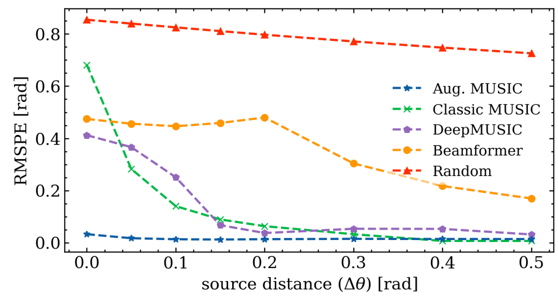

Fig. 5: DoA estimation of $d = 2$ closely spaced sources.

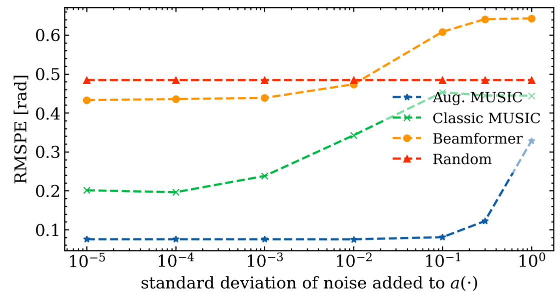

Fig. 6: DoA estimation with mismatch in the array geometry.

To compare the resolution of the algorithms, Fig. 5 shows the RMSPE for localizing $d = 2$ non-coherent signals, which are located close together with a $\Delta\theta$ distance from each other. The MB MUSIC algorithm is shown to collapse when the angular difference approaches $\Delta\theta \approx 0.1$ radians, while the deep augmented MUSIC demonstrates a constant low error for all $\Delta\theta$, indicating its improved resolution.

Finally, Fig. 6 depicts the RMSPE achieved when each element of the steering vector $\mathbf{a}(\cdot)$ is corrupted with zero-mean Gaussian noise, leading to a mismatch from the values used to compute the spatial spectrum. The deep augmented MUSIC is shown to learn to overcome such mismatches in the array geometry from the data. This indicates the improved robustness to array mismatch of the deep augmented MUSIC compared to its MB counterpart.

5. CONCLUSIONS

We presented a hybrid MB/DD implementation of the MUSIC algorithm for DoA estimation. The proposed deep augmented MUSIC was shown to mitigate some of the limitations and drawbacks of the classic method. Additionally, deep augmented MUSIC is adaptable to various scenarios and robust towards severe mismatches in the array geometry. The proposed hybrid MB/DD approach provides a viable alternative in both low and high snapshot domains, and shows a remarkable resolution capability compared to both MB and DD benchmarks.

6. REFERENCES

[1] Jeffrey Foutz, Andreas Spanias, and Mahesh K Banavar, "Narrowband direction of arrival estimation for antenna arrays," *Synthesis Lectures on Antennas*, vol. 3, no. 1, pp. 1–76, 2008.

[2] Maurice S Bartlett, "Smoothing periodograms from time-series with continuous spectra," *Nature*, vol. 161, no. 4096, pp. 686–687, 1948.

[3] Jack Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[4] Ralph Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[5] Soumitro Chakrabarty and Emanuël AP Habets, "Multi-speaker DOA estimation using deep convolutional networks trained with noise signals," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 8–21, 2019.

[6] Zhang-Meng Liu, Chenwei Zhang, and S Yu Philip, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 7315–7327, 2018.

[7] Nir Shlezinger, Jay Whang, Yonina C Eldar, and Alexandros G Dimakis, "Model-based deep learning," *arXiv preprint arXiv:2012.08405*, 2020.

[8] Ahmet M Elbir, "DeepMUSIC: Multiple signal classification via deep learning," *IEEE Sensors Letters*, vol. 4, no. 4, pp. 1–4, 2020.

[9] Andreas Barthelme and Wolfgang Utschick, "DoA estimation using neural network-based covariance matrix reconstruction," *IEEE Signal Process. Lett.*, 2021.

[10] Hamid Krim and John G Proakis, "Smoothed eigenspace-based parameter estimation," *Automatica*, vol. 30, no. 1, pp. 27–38, 1994, Special issue on statistical signal processing and control.

[11] Oren Solomon, Regev Cohen, Yi Zhang, Yi Yang, Qiong He, Jianwen Luo, Ruud JG van Sloun, and Yonina C Eldar, "Deep unfolded robust PCA with application to clutter suppression in ultrasound," *IEEE Trans. Med. Imag.*, 2019.

[12] Tirza Routtenberg and Joseph Tabrikian, "Bayesian parameter estimation using periodic cost functions," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1229–1240, 2011.

[13] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[14] Xavier Glorot and Yoshua Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, May 2010, vol. 9, pp. 249–256.