DEEP LEARNING-BASED COMPRESSIVE SAMPLING OPTIMIZATION IN MASSIVE MIMO SYSTEMS

Saidur R. Pavel [1], *Yimin D. Zhang* [1], *Maria S. Greco* [2], *and Fulvio Gini* [2]

[1] Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA
[2] Department of Information Engineering, University of Pisa, 56122 Pisa, Italy

ABSTRACT

In this paper, we develop a deep learning framework to optimize the compressive sampling matrix in a massive multiple-input multiple-output (MIMO) system. The optimized compressive sampling matrix is utilized to project high-dimensional data received at the massive MIMO system into a lower-dimensional space so that the directions of arrival and other signal parameters can be efficiently obtained with a reduced hardware complexity. The proposed deep learning approach for optimizing the compressive measurement matrix increases its robustness and generalizability.

***Index Terms*—** Massive MIMO, direction of arrival estimation, compressive sampling matrix, deep learning.

1. INTRODUCTION

Massive multiple-input multiple-output (MIMO) is considered a promising technology for the next-generation wireless communications and beyond. By equipping a high number of antennas, a MIMO system can provide enhanced system capacity, energy efficiency, security, and robustness [1–6]. Moreover, the highly directional beams offered by the massive MIMO system can also combat the severe propagation attenuation of millimeter-wave (mmWave) channels [7**?** , 8]. Despite the aforementioned benefits, dedicating a separate radio frequency (RF) frontend circuit to each antenna to obtain the baseband signals for subsequent processing requires prohibitively high hardware cost, power consumption, and computational complexity. Consequently, a hybrid analog-digital processing strategy, which reduces the number of RF frontend circuits connected to all antenna elements through a network of analog phase shifter, provides an attractive solution that renders a hybrid beamforming strategy with both hardware and computational efficiencies.

To achieve this objective, a compressive sampling matrix is optimized to reduce the dimensionality of the array received signal. Although using random measurement matrices is a popular option for this purpose, such unoptimized processing results in information loss and performance degradation [9]. In [10, 11], an information-theoretic framework is developed to optimize the compressive sampling matrix by maximizing the mutual information between signal direction of arrival (DOA) and the compressed measurements of the array received signal based on a coarse prior distribution. However, the coarse prior information may not always be available in practice, thereby limiting the applicability of this strategy. To address this issue, [12] proposed an iterative optimization strategy in which the normalized spatial spectrum estimate in each iteration is used as the prior information for the subsequent iteration, thereby eliminating the requirement for a coarse prior.

Optimizing the compressive measurement matrix in a sequential adaptive setting may result superior performance compared to the non-adaptive schemes [13, 14]. Solving such optimization problems using, for example, the projected gradient descent algorithm or a simplified version of the projected coordinate descent algorithm, to obtain the desired compressive sampling matrix is, nevertheless, computationally intensive. Alternatively, codebook-based methods, such as the hierarchical codebook developed in [14] and hierarchical Posterior Matching (hiePM) strategy developed in [4], can reduce the computational burden. However, the performance of the codebook based methods is highly dependent on the quality of the codebooks and may be inferior to that of the codebook-free methods.

Recently, deep learning techniques become attractive for solving complex optimization problems in an efficient manner. They have been applied in various wireless communication and signal processing contexts, such as massive MIMO beamforming [15, 16], intelligent reflecting surface [17, 18], and DOA estimation [19–21]. A deep learning-based approach to design adaptive beamformers for mmWave initial alignment is discussed in [22], where only single-antenna single-RF-chain scenarios are considered. In a more general massive MIMO system, however, multiple signals with multiple RF chains must be considered. The consideration of multiple signals is particularly important in MIMO radar and integrated sensing and communication (ISAC) applications as well as in an interference environment where the number of signal arrivals are unknown [23–27].

In this paper, we develop a deep learning-based sequential strategy to optimize the compressive measurement matrix in a multi-signal scenario. Because the neural networks can be trained offline, adopting deep learning to solve such sequential problems is computationally efficient. Moreover, compared to existing approaches [10, 11], such approaches provide a more generalized solution because the optimization is aided by a large number of training data.

Notations: We use lower-case and upper-case bold characters to denote vectors and matrices, respectively. In particular, $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ respectively represent the transpose and Hermitian operations of a matrix or vector. Notations $\div$ and $(.)^{②}$ denote element-wise division and element-wise square operations, respectively. In addition, $\mathrm{vec}(\cdot)$ vectorizes a matrix and $\det(\cdot)$ represents the determinant operator. $\mathbb{E}[\cdot]$ stands for the statistical expectation operator. $\mathcal{R}$ and $\mathcal{I}$ respectively extract the real and imaginary parts of a complex entry, and $\mathbb{C}^{M\times N}$ denotes the $M \times N$ complex space.

2. SIGNAL MODEL

Consider D uncorrelated far-field signals impinging on a massive MIMO system equipped with N receive antennas from directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_D]^{\mathrm{T}}$. The baseband array received signal vector is modeled as

$$\boldsymbol{x}(t) = \sum_{d=1}^{D} \boldsymbol{a}(\theta_D) s_d(t) + \boldsymbol{n}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{n}(t), \qquad (1)$$

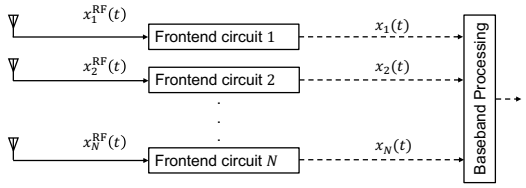


Fig. 1: Block diagram of an antenna array.

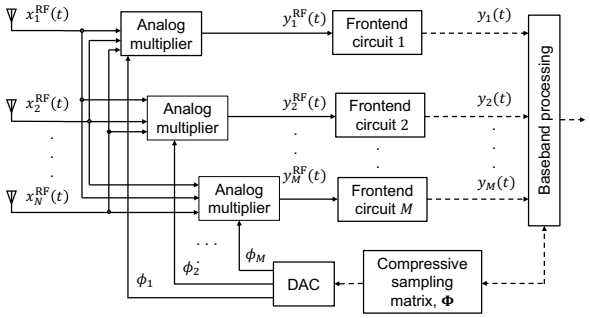


Fig. 2: Block diagram of the system exploiting compression.

where $\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \cdots, \boldsymbol{a}(\theta_D)] \in \mathbb{C}^{N\times D}$ is the array manifold matrix with dth column $\boldsymbol{a}(\theta_d)$ representing the steering vector corresponding to θ_d, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_D(t)]^{\mathrm{T}} \in \mathbb{C}^D$ denotes the signal waveform vector, and $\boldsymbol{n}(t) \sim \mathcal{CN}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}_N)$ denotes the zero-mean additive white Gaussian noise (AWGN) with noise power σ_n^2.

As shown in Fig. 1, separate RF frontend circuits are allocated to each antenna to obtain the baseband signal $\boldsymbol{x}(t)$ from the array received signal $\boldsymbol{x}^{\mathrm{RF}}(t)$. This scenario, however, is impractical for large-scale antenna arrays due to its high hardware requirements. To address this issue, we project the array received signal vector with dimension N onto a lower-dimensional space of dimension M with $M \ll N$. Each output channel is associate with a measurement kernel, described as a row vector $\{\boldsymbol{\phi}_m, m = 1, \cdots, M\} \in \mathbb{C}^{1\times N}$ as depicted in Fig. 2. Note in Fig. 2 that the solid lines and dashed lines denote the analog and digital signal flows, respectively.

After stacking the measurement kernels, the compressive sampling matrix is obtained as $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1^{\mathrm{T}}, \cdots, \boldsymbol{\phi}_M^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{M\times N}$. The compressive sampling matrix $\boldsymbol{\Phi}$ is used to obtain an M-dimensional compressed measurement vector of the N-dimensional array received signal $\boldsymbol{x}$ as

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}\boldsymbol{x}(t) = \boldsymbol{\Phi}\boldsymbol{A}(\theta)\boldsymbol{s}(t) + \boldsymbol{\Phi}\boldsymbol{n}(t), \tag{2}$$

where $\boldsymbol{\Phi}\boldsymbol{A}(\theta) \in \mathbb{C}^{M\times D}$ represents the compressed array manifold matrix. The main objective described in this paper is the optimization of the compressive sampling matrix $\boldsymbol{\Phi}$.

3. PROBABILISTIC SIGNAL MODEL

In this section, we summarize the probabilistic signal model described in [10, 12]. We treat the signal DOA θ as a random variable with *a priori* distribution $f(\theta)$. The probability density function (pdf) of the compressed measurement $\boldsymbol{y}$ can be expressed as

$$f(\boldsymbol{y}) = \mathbb{E}_\theta\{f(\boldsymbol{y}|\theta)\} = \int_{\theta\in\Theta} f(\boldsymbol{y}|\theta) f(\theta) \mathrm{d}\theta, \tag{3}$$

where Θ is the angular observation region. We discretize the pdf $f(\theta)$ into K angular bins, each with a width of $\Delta\bar{\theta}$. Then, the distribution of the compressed measurement $\boldsymbol{y}$ can be approximated as

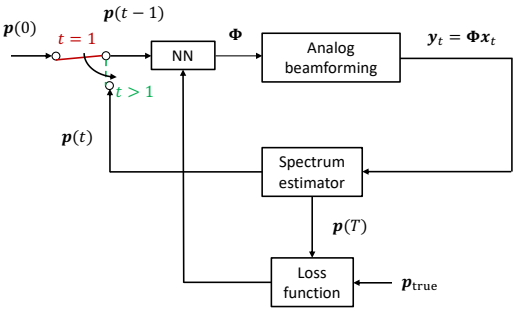


Fig. 3: Neural network framework for optimizing $\boldsymbol{\Phi}$.

$$f(\boldsymbol{y}) \approx \sum_{k\in\mathcal{K}} p_k f(\boldsymbol{y}|\theta_k), \tag{4}$$

where $p_k = f(\bar{\theta}_k)\Delta\bar{\theta}$ with $\sum_{k\in\mathcal{K}} p_k = 1$ denotes the probability of the kth angular bin, where $\bar{\theta}_k$ is the nominal angle at the kth bin and $\mathcal{K} = \{1, 2, \cdots, K\}$ is the set of the indices of the available angular bins. The signal arriving in a particular angular bin can be modeled as a zero-mean complex Gaussian random variable $s(t) \sim \mathcal{CN}(0, \sigma_s^2)$. As a result, the compressed measurement $\boldsymbol{y}$ forms a Gaussian mixture of K components. Then, the compressed measurement vector $\boldsymbol{y}$ for a signal impinging with a nominal DOA $\bar{\theta}_k$ can be expressed as

$$\boldsymbol{y}|_{\theta=\bar{\theta}_k} = \boldsymbol{\Phi}\left(\boldsymbol{a}(\bar{\theta}_k)s(t) + \boldsymbol{n}(t)\right) \tag{5}$$

with conditional pdf

$$f(\boldsymbol{y}|\bar{\theta}_k) = \frac{1}{\pi^M \det(\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k})} e^{-\boldsymbol{y}^{\mathrm{H}} \boldsymbol{C}^{-1}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k} \boldsymbol{y}}, \tag{6}$$

where $\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k} = \boldsymbol{\Phi}(\sigma_s^2 \boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k) + \sigma_n^2 \boldsymbol{I}_N)\boldsymbol{\Phi}^{\mathrm{H}}$ is the covariance matrix of the compressed measurement $\boldsymbol{y}$ for a particular DOA $\bar{\theta}_k$ and σ_s^2 is the estimated signal power.

4. DEEP LEARNING-BASED OPTIMIZATION OF THE COMPRESSIVE SAMPLING MATRIX

In this section, we describe the proposed deep learning framework for optimizing the compressive measurement matrix $\boldsymbol{\Phi}$, as shown in Fig. 3. Because no prior information about the source distribution is assumed, we begin the optimization procedure by considering a uniform prior at time $t = 0$. The compressive sampling matrix $\boldsymbol{\Phi}$ at time t can be considered as a function of the posterior distribution of the spatial spectrum and the compressive sample matrices of the past observations, given as

$$\boldsymbol{\Phi}(t) = \mathcal{F}\left(\boldsymbol{p}(1:t-1), \boldsymbol{\Phi}(1:t-1)\right), \tag{7}$$

where vector $\boldsymbol{p}(t)$ denotes the posterior of θ at time t and $\mathcal{F}$ is a function that maps the current compressive sampling matrix from the past posteriors $\boldsymbol{p}(1:t-1)$ and the past compressive sampling matrices $\boldsymbol{\Phi}(1:t-1)$. As pointed out in [28], instead of using all of the past observations, the posterior of the DOAs at time t, $\boldsymbol{p}(t-1)$, provides sufficient statistics for the design of $\boldsymbol{\Phi}(t)$, i.e., $\boldsymbol{\Phi}(t) = \hat{\mathcal{F}}(\boldsymbol{p}(t-1))$ with $\hat{\mathcal{F}}$ being a mapping function.

We design a neural network to realize the function $\hat{\mathcal{F}}$. We consider an active learning framework composed of an L-layer fully

connected (FC) network that iteratively updates the posterior to obtain the best posterior and accurately resolve the signal DOAs. There might other possible neural network architectures available for this problem, but in this paper we focus on the FC network due to its implementation simplicity. As such, we consider an L-layer FC network to obtain the compressive sampling matrix $\boldsymbol{\Phi}(t)$ based on the DOA distribution $\boldsymbol{p}(t-1)$ from the previous time frame. Note that, as we will see, $\boldsymbol{p}(t-1)$ is obtained based on $\boldsymbol{y}(t-1)$ and thus reflects the past compressive sampling matrix $\boldsymbol{\Phi}(t-1)$.

Define $\tilde{t} = (t-1)/T$ as the normalized time index, where T is the number of snapshots. We use the posterior $\boldsymbol{p}(t-1)$ and the normalized time index $\tilde{t}$ as the input to the neural network at time t, i.e., $\boldsymbol{v}(t-1) = [\boldsymbol{p}^{\mathrm{T}}(t-1) \;\; \tilde{t}]^{\mathrm{T}}$. For a total of B observations in a particular batch of the training data, the complete training dataset $\boldsymbol{G}(t-1)$ is formed by concatenating vectors $\boldsymbol{v}_b(t)$ corresponding to the observations $b \in \{1, 2, \cdots, B\}$ as $\boldsymbol{G}(t-1) = [\boldsymbol{v}_1^{\mathrm{T}}(t-1), \boldsymbol{v}_2^{\mathrm{T}}(t-1), \cdots, \boldsymbol{v}_B^{\mathrm{T}}(t-1)]$. The neural network output provides the compressive sampling matrix for the next measurement as

$$\tilde{\boldsymbol{\Phi}}(t) = \mathcal{A}_L\left(\boldsymbol{W}_L \mathcal{A}_{L-1}(\cdots \mathcal{A}_1(\boldsymbol{W}_1 \boldsymbol{G}(t-1) + \boldsymbol{b}_1)\cdots) + \boldsymbol{b}_L\right), \tag{8}$$

where $\{\boldsymbol{W}_l, \boldsymbol{b}_l, \mathcal{A}_l\}_{l=1}^{L}$ are the weights, biases, and nonlinear activation function corresponding to the lth layer, respectively. $\tilde{\boldsymbol{\Phi}}(t)$ is an augmented matrix that denotes the real-valued representation of the complex-valued compressive sampling matrix at time t, i.e., $\tilde{\boldsymbol{\Phi}}(t) = [\mathcal{R}(\boldsymbol{\Phi}(t)) \;\; \mathcal{I}(\boldsymbol{\Phi}(t))]$. The required $\boldsymbol{\Phi}(t)$ can then be extracted from $\tilde{\boldsymbol{\Phi}}$, where the real and imaginary parts of $\boldsymbol{\Phi}(t)$ respectively correspond to the left and right halves of $\tilde{\boldsymbol{\Phi}}(t)$.

The measurement kernels $\boldsymbol{\phi}_m, m = 1, 2, \cdots, M$, are generally implemented using a series of phase shifters. As such, it is desirable for the compressive measurement matrix to satisfy a constant modulus constraint. Toward this end, we set the activation function of the last layer as

$$\begin{aligned} \mathcal{A}_L(\mathcal{R}(\boldsymbol{\Phi})) &= \left[\mathcal{R}(\boldsymbol{\Phi}) \div \sqrt{\mathcal{R}(\boldsymbol{\Phi})^{②} + \mathcal{I}(\boldsymbol{\Phi})^{②}}\right], \\ \mathcal{A}_L(\mathcal{I}(\boldsymbol{\Phi})) &= \left[\mathcal{I}(\boldsymbol{\Phi}) \div \sqrt{\mathcal{R}(\boldsymbol{\Phi})^{②} + \mathcal{I}(\boldsymbol{\Phi})^{②}}\right]. \end{aligned} \tag{9}$$

Using the neural network output $\boldsymbol{\Phi}(t)$, we then form an analog beamformer to obtain compressed measurements $\boldsymbol{y}(t)$ from $\boldsymbol{x}(t)$ at time t. We use the minimum variance distortionless response (MVDR) spatial spectrum estimator based on the compressed measurement vector to find the spatial spectrum as

$$P_{\text{CS-MVDR}}^{(t)}(\theta) = \frac{1}{N} \frac{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{\Phi}^{\mathrm{H}}(t)\boldsymbol{\Phi}(t)\boldsymbol{a}(\theta)}{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{\Phi}^{\mathrm{H}}(t)\hat{\boldsymbol{R}}_{\boldsymbol{yy}}^{-1}(t)\boldsymbol{\Phi}(t)\boldsymbol{a}(\theta)}, \tag{10}$$

where $\hat{\boldsymbol{R}}_{yy}(t)$ is the sample covariance matrix of $\boldsymbol{y}$ estimated at time t, which is computed based on the current and past snapshots as

$$\hat{\boldsymbol{R}}_{\boldsymbol{yy}}(t) = \beta \hat{\boldsymbol{R}}_{\boldsymbol{yy}}(t-1) + \boldsymbol{y}(t)\boldsymbol{y}^{\mathrm{H}}(t), \tag{11}$$

where $\hat{\boldsymbol{R}} = \boldsymbol{0}$ at time $t = 0$ and β is a forgetting factor. To avoid ill conditioned matrix inversion of $\hat{\boldsymbol{R}}_{\boldsymbol{yy}}$, the computation of the spatial spectrum $P_{\text{CS-MVDR}}$ and update of the posterior take place only when $M \le t < T$.

The normalized spatial spectrum can be considered as the posterior distribution of the DOAs at time t, i.e.,

$$\boldsymbol{p}(t) = \frac{[P_{\text{CS-MVDR}}^{(t)}(\theta_1), \cdots, P_{\text{CS-MVDR}}^{(t)}(\theta_K)]}{\sum_{k=1}^{K} P_{\text{CS-MVDR}}^{(t)}(\theta_k)} \tag{12}$$

for $M \le t < T$. The obtained posterior distribution $\boldsymbol{p}(t)$ is then fed to the neural network again for sequential optimization of $\boldsymbol{\Phi}$.

During the iteration through time samples, we keep the parameters of the neural network unchanged to yield more scalable and faster training procedure. Once the iterations through all time samples are completed, we update the neural network parameters by minimizing a suitable loss function. We consider the problem as a regression problem and use the mean square error (MSE) loss function between the estimated DOA distribution $\boldsymbol{p}_i(T)$ and the true distribution $\boldsymbol{p}_{i_{\text{true}}}$ for ith batch, expressed as

$$\text{Loss} = \frac{1}{BK} \sum_{i=1}^{B} \left\| \boldsymbol{p}_i(T) - \boldsymbol{p}_{i_{\text{true}}} \right\|^2. \tag{13}$$

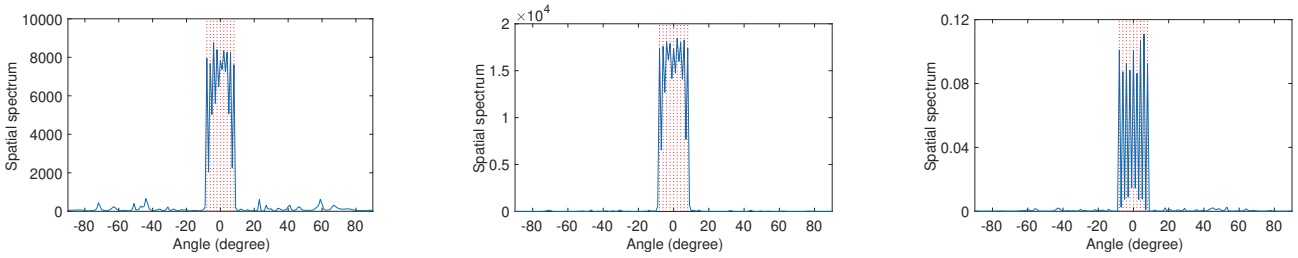
5. SIMULATION REULTS

Consider a massive MIMO system consisting of $N = 50$ receive antennas which are arranged in a uniform linear fashion separated by half wavelength. 9 far-field uncorrelated sources impinge on the antenna array. We choose the compression ratio $N/M = 5$, which results the dimension of the compressed measurements to be $M = 10$. The pdf of the DOAs are discretized with an width of $\Delta\bar{\theta} = 1^\circ$, rendering 181 components in the Gaussian mixture model. The number of snapshots is $T = 30$.

The proposed neural network consists of 4 fully connected layers and each layer contains 500 nodes. The number of layers and the nodes are chosen experimentally by considering the generalization capability, overfitting risk, and optimization complexity of the network. We introduce a dropout regularizer at each fully connected layer with a rate of 0.3, which randomly discard a subset of neural network nodes to reduce potential overfitting of the training data. We generate $10,000$ scenarios for the training dataset, where each scenario consists of 9 sources uniformly sampled from the discrete grid of $[-90^\circ, 90^\circ]$. In a similar fashion a test dataset consisting of 1000 samples are generated. The input signal-to-noise-ratio (SNR) is chosen randomly between 0 dB to 20 dB from a uniform distribution to train the network. The forgetting factor is set to $\beta = 0.9$.

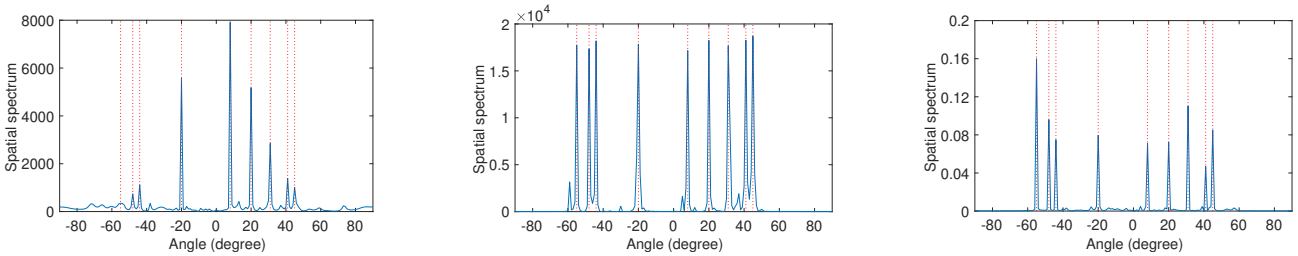
Assuming no prior knowledge of the DOA distribution, the training procedure of the neural network begins by considering the prior information of each scenario for a particular batch to be uniformly distributed. This prior information along with the time samples will act as the input to the network. We also prepare the label of network by making the nominal probability $1/D$ in the grid points where the sources are actually present, whereas the remaining grid points are set to zero. The neural network is trained to learn the mapping $\hat{\mathcal{F}}$ from the prior to the compressive sampling matrix by minimizing the MSE loss function. The Adam optimizer is used with an learning rate progressively decreasing from 0.1 to 0.001. We also employ minibatch training with a batch size of 64, and 200 epochs are completed to train the network.

We compare the performance of the proposed method to those described in [11, 12]. In particular, reference [11] assumes that the prior distribution of the DOAs follows a Gaussian distribution. A presumed Gaussian distribution with mean 0° and variance $(5^\circ)^2$ is considered for this method. This method provides a good performance only when the actual signals closely follow the presumed parameters. To examine the performance of the proposed method and those described in [11, 12] for different situations, we consider two test cases, both with 9 signals. In the first case, the signals closely follows the presumed distribution, where the DOAs are -8°, -6°, -4°, -2°, 0°, 2°, 4°, 6°, and 8°. On the other hand, in the second



(a) Method in [11] (b) Method in [12] (c) Proposed method

Fig. 4: Estimated spatial spectra for test case 1.



(a) Method in [11] (b) Method in [12] (c) Proposed method

Fig. 5: Estimated spatial spectra for test case 2.

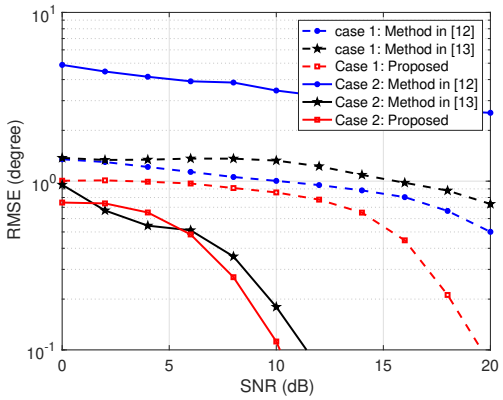


Fig. 6: Comparison of the RMSE performance.

case, the signal arrivals significantly deviate from the presumed distribution where the DOAs are $-55°$, $-48°$, $-44°$, $-20°$, $8°$, $20°$, $31°$, $41°$, and $45°$.

As shown in Fig. 4a, for test case 1, the method in [11] resolves all of the sources successfully. On the other hand, for the second case, as the impinging signals deviate from the assumed spatial distributions, the performance of this method degrades. As depicted in Fig. 5a, the signals that are away from the expected mean are highly attenuated.

The method described in [12] eliminates the requirement of the prior knowledge by iteratively updating the prior information. Beginning by a uniform prior, this method estimates the spatial spectrum of the impinging signals and uses this information as the updated prior for the subsequent iterations. As shown in Figs. 4b and 5b, this method provides satisfactory performance in both cases. Note that this method requires real-time optimization with a high implementation complexity.

Figs. 4c and 5c depict the performance of the proposed deep learning-based method in the considered two cases. In both cases, the proposed method provides superior performance with enhanced signal resolution compared to other two methods. Once the network completes the training, which can be carried out offline, this method can quickly resolve the DOAs. Another advantage of the proposed deep learning-based optimization is its generalization capability. Unlike the conventional approach, where separately estimate every realization of the posterior from scratch, the trained neural network can generate an optimal compressive sampling matrix for a variety of scenarios, provided that a large training dataset is available.

Fig. 6 compares the performance of the proposed method with the methods described in [11, 12] in terms of the root mean squared error (RMSE), defined as

$$\text{RMSE} = \sqrt{\frac{1}{QD}\sum_{q=1}^{Q}\sum_{d=1}^{D}(\hat{\theta}_{q,d} - \theta_d)^2} \tag{14}$$

where Q is the number of trials and $\hat{\theta}_{q,d}$ is estimated DOA for the dth source of the qth Monte-Carlo trial.

For each input SNR value, 300 Monte Carlo trials are used to compute the RMSE. The dotted lines denote the results of test case 1, where the proposed method provides superior performance compared to the others. For test case 2 as depicted with the solid lines, the method in [11] does not do well since the actual signal arrivals do not match the assumed prior distribution. The proposed method and the method developed in [12] provide low RMSE results.

6. CONCLUSION

In this paper, we developed a neural network-based framework to optimize a compressive sampling matrix in a massive MIMO system, which efficiently projects high-dimensional data into a lower-dimensional space. The optimal compressive sampling matrix obtained from the neural network allows user signal parameters to be obtained from compressed measurements, reducing the required RF frontend circuit. Our proposed method provides more accurate, robust, and general solutions compared to existing literature.

7. REFERENCES

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2012.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[4] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.

[5] F. Jiang, J. Chen, A. L. Swindlehurst, and J. A. López-Salcedo, "Massive MIMO for wireless sensing with a coherent multiple access channel," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3005–3017, 2015.

[6] Y. Gu and Y. D. Zhang, "Information-theoretic pilot design for downlink channel estimation in FDD massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2334–2346, 2019.

[7] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, 2014.

[8] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, 2017.

[9] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, "Analysis of Fisher information and the Cramer-Rao bound for nonlinear parameter estimation after compressed sensing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2013, pp. 6630–6634.

[10] Y. Gu, Y. D. Zhang, and N. A. Goodman, "Optimized compressive sensing-based direction-of-arrival estimation in massive MIMO," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2017, pp. 3181–3185.

[11] Y. Gu and Y. D. Zhang, "Compressive sampling optimization for user signal parameter estimation in massive MIMO systems," *Digital Signal Process.*, vol. 94, pp. 105–113, 2019.

[12] Y. D. Zhang, "Iterative learning for optimized compressive measurements in massive MIMO systems," in *Proc. IEEE Radar Conf.*, 2022, pp. 1–5.

[13] V. Nakos, X. Shi, D. P. Woodruff, and H. Zhang, "Improved algorithms for adaptive compressed sensing," in *Proc. Int. Colloq. Automata, Lang., Program. (ICALP)*, 2018, pp. 90:1–14.

[14] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6222–6235, 2011.

[15] Y. Yang, S. Zhang, F. Gao, C. Xu, J. Ma, and O. A. Dobre, "Deep learning based antenna selection for channel extrapolation in FDD massive MIMO," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2020, pp. 182–187.

[16] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Trans. Vehi. Tech.*, vol. 69, no. 1, pp. 1065–1069, 2019.

[17] S. Zhang, S. Zhang, F. Gao, J. Ma, and O. A. Dobre, "Deep learning optimized sparse antenna activation for reconfigurable intelligent surface assisted communication," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6691–6705, 2021.

[18] T. Jiang, H. V. Cheng, and W. Yu, "Learning to reflect and to beamform for intelligent reflecting surface with implicit channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1931–1945, 2021.

[19] L. Wu, Z.-M. Liu, and Z.-T. Huang, "Deep convolution network for direction of arrival estimation with sparse prior," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1688–1692, 2019.

[20] S. R. Pavel, M. W. T. Chowdhury, Y. D. Zhang, D. Shen, and G. Chen, "Machine learning-based direction-of-arrival estimation exploiting distributed sparse arrays," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, 2021, pp. 241–245.

[21] S. R. Pavel and Y. D. Zhang, "Neural network approach to iterative optimization of compressive measurement matrix in massive MIMO system," in *Proc. IEEE Sensor Array and Multich. Signal Process. (SAM) Workshop*, June 2022, pp. 171–175.

[22] F. Sohrabi, Z. Chen, and W. Yu, "Deep active learning approach to adaptive beamforming for mmWave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2347–2360, 2021.

[23] A. Hassanien, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Signaling strategies for dual-function radar communications: An overview," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 31, no. 10, pp. 36–45, 2016.

[24] S. Fortunati, L. Sanguinetti, F. Gini, M. S. Greco, and B. Himed, "Massive MIMO radar for target detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 859–871, 2020.

[25] A. M. Ahmed, A. A. Ahmad, S. Fortunati, A. Sezgin, M. S. Greco, and F. Gini, "Reinforcement learning based beamforming for massive MIMO radar multi-target detection," *arXiv preprint arXiv:2005.04708*, 2020.

[26] S. Sun and Y. D. Zhang, "4D automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 4, pp. 879–891, 2021.

[27] F. Lisi, S. Fortunati, M. S. Greco, and F. Gini, "Enhancement of a state-of-the-art RL-based detection algorithm for massive MIMO radars," *IEEE Trans. Aerosp. Electron. Syst.*, in press, doi:10.1109/TAES.2022.3168033.

[28] S.-E. Chiu, N. Ronquillo, and T. Javidi, "Active learning and CSI acquisition for mmWave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2474–2489, 2019.