NEURAL NETWORK APPROACH TO ITERATIVE OPTIMIZATION OF COMPRESSIVE MEASUREMENT MATRIX IN MASSIVE MIMO SYSTEM

Saidur R. Pavel and Yimin D. Zhang

Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA

ABSTRACT

Massive multiple-input multiple-output (MIMO) systems equipped with a high number of antenna elements are becoming one of the most exciting emerging technologies in next-generation wireless communication systems. The use of a high number of antennas makes it challenging to implement separate radio frequency front-end circuits for each antenna due to hardware cost and power consumption restrictions. To make massive MIMO feasible, an efficient solution is to perform hybrid beamforming that compresses the received signals at the MIMO receiver before the signals are digitized. We have proposed optimized compressive measurement by maximizing the mutual information between the compressed measurement and the signal directions-of-arrival (DOAs) by utilizing a coarse *a priori* probability distribution of the signal DOAs. Despite its superior performance to effectively reduce the number of required front-end circuits, the requirement of a coarse *a priori* distribution of signal DOAs makes it difficult to apply in some situations where such information is unavailable. In this paper, we propose a data-driven approach based on neural network to iteratively update the signal DOA distribution. The neural network estimate signal DOA spectrum, which is then fed back to refine the prior information, thereby making this method practical even in the situations where no prior information of the signal DOAs is available.

Index Terms— Massive MIMO, compressive measurement, neural network, iterative learning.

1. INTRODUCTION

Massive multiple-input multiple-output (MIMO) technology is a key enabler in next-generation wireless communications. Massive MIMO systems provide spectrum-efficient, energy-efficient, robust, and secure solutions due to the large number of antenna elements deployed in the base station [1–7]. The utilization of a large number of antennas enables the systems to serve multiple users with a high array gain and reduced multiuser interference. Massive MIMO systems address the issues of high propagation attenuation of millimeter-wave communications and make effective bandwidth utilization [8–10]. Massive MIMO systems are also becoming increasingly important in radar sensing [11–13].

Despite all these offerings of massive MIMO systems, the high number of antenna elements require considerably high hardware cost and power consumption. In particular, it is challenging and impractical to dedicate a separate radio frequency (RF) front-end chain and a high-resolution analog-to-digital converter (ADC) to each antenna element. To make massive MIMO practicable, one solution is to use hybrid analog-digital beamforming, in which a reduced number of RF chains and ADCs are used. Toward this end, compressive sampling technique can be used to project a high-dimensional array signal into a low-dimensional manifold by exploiting the sparsity nature of signals. A common way to perform compressive sampling is to use random measurement matrices [14–16]. Although adopting random measurement matrices is robust as it guarantees the sparse signal recovery with a high probability [17–19], such approaches generally cause severe information losses [20].

In [21, 22], a coarse probability distribution of the user directions-of-arrival (DOAs) is exploited to optimize the compressive measurement matrix. By maximizing the mutual information between the compressed measurements and the signal DOA, this approach effectively reduces the number of front-end circuits with negligible performance loss. In some applications, however, the required knowledge of a coarse *a priori* distribution of the signal DOAs may not be available. To optimize the compressive measurement matrix without such *a priori* information, an iterative learning scheme is proposed in [23] which uses the estimated DOA spectrum as the prior information for the subsequent iteration.

In this paper, we propose a data-driven approach based on neural network to estimate the signal DOAs from the compressed measurement and feed this information from the neural network output as *a priori* knowledge back to the system to further optimize the compressive measurement matrix. Adopting data-driven approach is advantageous compared to the minimum variance distortionless response (MVDR) spectrum estimator used in [23]. First, model-driven approaches, such as the MVDR spectrum estimator, formulate a parametric mapping from the signal direction to the compressed measurement of the array output. As such, it does not address the information loss incurred in the compression process. In the proposed data-driven approach, on the other hand, a nonlinear relationship is optimized to minimize a loss function that relates the compressed measurements and the actual DOAs in the training process based on a large number of training data, thereby achieving improved performance. Second, data-driven approaches further provide resilient solutions in the presence of impaired array model [24, 25]. When the array model is impaired, such as imperfect calibration or knowledge of the sensor gain, phase, position, and mutual coupling, the forward mapping from signal DOAs to array input signals becomes complicated. Without a precise formulation of these impairements, parametric approaches like the MVDR spectrum estimator degrade in performance. On the other hand, the proposed data-driven approach does not rely on any presumptions but learns the impairment from the data, thus resulting in a robust optimization approach.

Notations: We use lower-case (upper-case) bold characters to denote vectors (matrices). 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. $\mathrm{triu}(\cdot)$ returns the upper triangular part of an matrix. In addition, $\mathrm{vec}(\cdot)$ vectorizes a matrix and $\mathrm{Tr}(\cdot)$ represents the trace operator, and $\mathrm{diag}(\cdot)$ forms a diagonal matrix. $\mathbb{E}_\theta[\cdot]$ denotes the statistical expectation with respect to θ. $\mathcal{R}$ and $\mathcal{I}$ respectively extract the real and imaginary parts of a complex entry. $\mathbb{C}^{M\times N}$ denotes the $M \times N$ complex space, and $|\boldsymbol{A}|$ denotes the determinant of matrix $\boldsymbol{A}$.

2. SIGNAL MODEL

Consider D uncorrelated signals that are impinging on a massive MIMO system equipped with N antennas from directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots \theta_D]^{\mathrm{T}}$. The baseband vector of the array received signal at time t is modeled as

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

Where $\bar{\boldsymbol{A}}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \cdots, \boldsymbol{a}(\theta_D)] \in \mathbb{C}^{N\times D}$ denotes the presumed array manifold matrix whose column $\boldsymbol{a}(\theta_d) \in \mathbb{C}^N$ represents the presumed steering vector of the dth user with DOA θ_d, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_D(t)]^{\mathrm{T}} \in \mathbb{C}^D$ represents the signal waveform vector, and $\boldsymbol{n}(t) \sim \mathcal{CN}(\boldsymbol{0}, \sigma_{\mathrm{n}}^2 \boldsymbol{I}_D)$ represents the zero-mean additive white Gaussian noise (AWGN) vector.

When an array sensor contains calibration errors, we describe the gain and phase errors of the nth sensor as $g_n = \alpha_n e^{\jmath\beta_n}$ for $n = 1, \cdots, N$, and denote $\boldsymbol{g} = [\alpha_1 e^{\jmath\beta_1}, \cdots, \alpha_N e^{\jmath\beta_N}]^{\mathrm{T}}$. Then, the actual array manifold $\boldsymbol{A}(\boldsymbol{\theta})$ is expressed as

$$\boldsymbol{A}(\boldsymbol{\theta}) = \mathrm{diag}(\boldsymbol{g})\bar{\boldsymbol{A}}(\boldsymbol{\theta}). \quad (2)$$

Denote $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1^{\mathrm{T}}, \boldsymbol{\phi}_2^{\mathrm{T}}, \cdots, \phi_M^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{M\times N}$ with $M \ll N$ as the compressive measurement matrix, where $\boldsymbol{\phi}_m \in \mathbb{C}^{1\times N}$ is a row vector denoting the measurement kernel for $m = 1, \cdots, M$. $\boldsymbol{\Phi}$ is designed to be row-orthonormal, i.e., $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}} = \boldsymbol{I}_M$, to keep the noise power unchanged after applying the compression. Projection of the high-dimensional signal vector $\boldsymbol{x}(t)$ into the mth measurement kernel $\boldsymbol{\phi}_m$ yields the mth compressive measurement $y_m(t) = \boldsymbol{\phi}_m \mathbf{x}(t)$. Stacking the M compressed measurement vector yields $\boldsymbol{y}(t) = [y_1(t), y_2(t), \cdots, y_M(t)]^{\mathrm{T}} \in \mathbb{C}^M$, which is given as

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}\boldsymbol{x}(t) = \boldsymbol{\Phi}\boldsymbol{A}(\theta)\boldsymbol{s}(t) + \boldsymbol{\Phi}\boldsymbol{n}(\boldsymbol{t}), \quad (3)$$

where $\boldsymbol{\Phi}\mathbf{A}(\theta) \in \mathbb{C}^{M\times D}$ represents the compressed array manifold with significantly reduced dimension.

This paper aims to optimize the compressive measurement matrix $\boldsymbol{\Phi}$ based on the maximization of mutual information $I(\boldsymbol{y}; \theta)$ between the compressed measurements $\boldsymbol{y}$ and the signal DOA θ without assuming any knowledge about the signal DOA distribution. The problem is formulated as

$$\begin{aligned} \max_{\boldsymbol{\Phi}} \quad & I(\boldsymbol{y}; \theta) \\ \text{s.t.} \quad & M = \mathcal{C}_R N, \end{aligned} \quad (4)$$

where $\mathcal{C}_R < 1$ denotes the desired compression ratio. We use a neural network to refine the prior information for further optimizing $\boldsymbol{\Phi}$ in an iterative manner.

3. OPTIMIZATION OF COMPRESSIVE MEASUREMENT MATRIX

In this section, we summarize the optimization approaches of the compressive measurement matrix developed in [21–23], whereas the proposed data-driven iterative learning is presented in Section 4.

3.1. Probabilistic Signal Model

Consider signal DOA θ as a random variable with a probability density function (PDF) $f(\theta)$. In [21, 22], it is assumed that coarse knowledge of $f(\theta)$ is available. In this case, according to the law of the total probability, the PDF of the compressed measurement vector $\boldsymbol{y}$ is expressed as

$$f(\boldsymbol{y}) = \mathbb{E}_\theta\{f(\boldsymbol{y}|\theta)\} = \int_{\theta\in\Theta} f(\boldsymbol{y}|\boldsymbol{\theta}) f(\theta) d\theta, \quad (5)$$

where Θ is the angular region of observation. We discretize the PDF $f(\theta)$ into K angular bins with an equal width of $\Delta\bar{\theta}$ so that the probability of the kth angular bin is given as $p_k \approx f(\bar{\theta}_k)\Delta\bar{\theta}$ with $\sum_{k\in\mathcal{K}} p_k = 1$, where $\bar{\theta}_k$ is the nominal angle of the kth angular bin and $\mathcal{K} = \{1, 2, \cdots, K\}$. As a result, the PDF of $\boldsymbol{y}$ becomes

$$f(\boldsymbol{y}) \approx \sum_{k\in\mathcal{K}} p_k f(\boldsymbol{y}|\bar{\theta}_k). \quad (6)$$

Considering a signal impinging from the kth angular bin with a nominal DOA $\bar{\theta}_k$, the compressed measurement vector is given as

$$\boldsymbol{y}|_{\theta=\bar{\theta}_k} = \boldsymbol{\Phi}(\boldsymbol{a}(\bar{\theta}_k)s(t) + \boldsymbol{n}(t)), \quad (7)$$

and the corresponding conditional PDF is

$$f(\boldsymbol{y}|\bar{\theta}_k) = \frac{1}{\pi^M |\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}|} e^{-\boldsymbol{y}^{\mathrm{H}} \boldsymbol{C}^{-1}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k} \boldsymbol{y}}, \quad (8)$$

where $\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k} = \boldsymbol{\Phi}(\sigma_s^2 \boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k) + \sigma_{\mathrm{n}}^2 \boldsymbol{I})\boldsymbol{\Phi}^{\mathrm{H}}$ is the covariance matrix of the compressed measurement $\boldsymbol{y}|_{\theta=\bar{\theta}_k}$ and σ_s^2 is the estimated signal power. Therefore, the PDF of $\boldsymbol{y}$ is approximated as a Gaussian mixture distribution.

3.2. Optimization of the Compressive Measurement Matrix

To optimize the compressive measurement matrix $\boldsymbol{\Phi}$ through the maximization of the mutual information between the compressed measurement vector $\boldsymbol{y}$ and the DOA θ, the gradient of the mutual information $I(\boldsymbol{y}; \theta)$ with respect to $\boldsymbol{\Phi}$ is obtained as

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y}; \theta) = \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}) - \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}|\theta), \quad (9)$$

where $\nabla_{\boldsymbol{\Phi}}$ represents the gradient operator with respect to $\boldsymbol{\Phi}$, $h(\boldsymbol{y}) = -\mathbb{E}_{\boldsymbol{y}}\{\log[f(\boldsymbol{y})]\}$ is the differential entropy of $\boldsymbol{y}$, and $h(\boldsymbol{y}|\theta) = -\mathbb{E}_{\boldsymbol{y},\theta}\{\log[f(\boldsymbol{y}|\theta)]\}$ is the conditional differential entropy of $\boldsymbol{y}$ given the signal DOA θ. An approximation of the mutual information gradient can be derived as [21, 22]:

$$\begin{aligned} &\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y}; \theta) \\ &\approx \frac{\sum\limits_{k\in\mathcal{K}} \frac{p_k}{\left|\frac{\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}}{\sigma_{\mathrm{n}}^2}\right|} \left[\frac{\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}}{\sigma_{\mathrm{n}}^2}\right]^{-1} \boldsymbol{\Phi}\left(\frac{\sigma_s^2}{\sigma_{\mathrm{n}}^2}\boldsymbol{a}(\bar{\theta}_k)\boldsymbol{a}^{\mathrm{H}}(\bar{\theta}_k) + \boldsymbol{I}\right)}{\sum\limits_{k\in\mathcal{K}} p_k \left|\frac{\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}}{\sigma_{\mathrm{n}}^2}\right|^{-1}} \\ &\quad - \sum_{k\in\mathcal{K}} \left[\frac{\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\boldsymbol{\theta}}_{\boldsymbol{k}}}}{\sigma_{\mathrm{n}}^2}\right]^{-1} \boldsymbol{\Phi}\left(\frac{\sigma_s^2}{\sigma_{\mathrm{n}}^2}\boldsymbol{a}\left(\bar{\theta}_k\right)\boldsymbol{a}^{\mathrm{H}}\left(\bar{\theta}_k\right) + \boldsymbol{I}\right), \end{aligned} \quad (10)$$

where $\sigma_s^2/\sigma_{\mathrm{n}}^2$ represents the estimated signal-to-noise ratio (SNR) of the input signal.

The obtained mutual information gradient $\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y}; \theta)$ is used to iteratively update the compressive measurement matrix $\boldsymbol{\Phi}$ in a gradient ascent manner, i.e.,

$$\boldsymbol{\Phi} \leftarrow \boldsymbol{\Phi} + \gamma \nabla_{\boldsymbol{\Phi}} I\left(\boldsymbol{y}; \theta\right), \quad (11)$$

where $\gamma > 0$ is the step size.

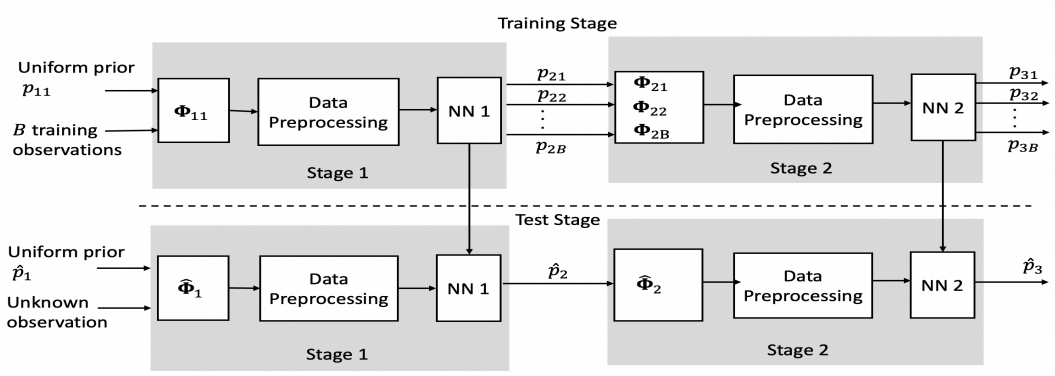


Fig. 1: Proposed iterative learning framework.

3.3. Iterative Learning Scheme

An MVDR estimator computes the spatial spectrum as [23]

$$P^{(i)}(\theta) = \frac{\boldsymbol{a}^{\mathrm{H}}(\theta)(\boldsymbol{\Phi}^{(i)})^{\mathrm{H}}\boldsymbol{\Phi}^{(i)}\boldsymbol{a}(\theta)}{\boldsymbol{a}^{\mathrm{H}}(\theta)(\boldsymbol{\Phi}^{(i)})^{\mathrm{H}}\left(\hat{\mathbf{R}}_{\boldsymbol{yy}}^{(i)}\right)^{-1}\boldsymbol{\Phi}^{(i)}\boldsymbol{a}(\theta)}, \tag{12}$$

where superscript $^{(i)}$ is added to indicate the ith iteration. When the signal DOA distribution is not available, We normalize the estimated spatial spectrum and use it as the prior distribution in the subsequently iteration, i.e., $\hat{p}_k^{(i+1)} = P^{(i)}(\theta_k)/\sum_{j\in\mathcal{K}} P^{(i)}(\theta_j)$. This learning process can be iterated. In the initial condition, when there is no prior information about the signal arrivals, a random compressive measurement matrix consisting of row-orthonormal Gaussian random elements is used.

4. NEURAL NETWORK-BASED ITERATIVE LEARNING

In this section, we propose a data-driven framework to estimate the signal DOAs from the compressed measurement and use the estimated spatial spectrum as the the prior information for the next iteration for further compression, as shown in Fig. 1. The subscript ij in Fig. 1 denotes the ith stage and the jth element. While the basic concept of iterative learning is similar to [23], the neural network output, in lieu of the MVDR estimates, is used to improve the accuracy and robustness against imperfections.

The proposed framework is divided into two stages. The first stage estimates the coarse probability distributions whereas the second stage refines the spatial spectrum. Note that a single distribution is used in the first stage for all observations and a single compressive measurement matrix $\mathbf{\Phi}$ is optimized. Once $\mathbf{\Phi}$ is optimized, it is kept fixed for this stage and the obtained compressed data are then preprocessed before they are fed into the neural network NN1. The neural network is trained by exploiting the actual DOAs as the label and produce B distributions which respectively correspond to different observations of the training dataset. For each of the distributions, a separate compressive measurement matrix is computed in the second stage. Thus, the compression performed in the second stage utilizes the specific prior corresponding to a particular observation. The preprocessed compressed data are then fed into the neural network NN2 to refine the spatial spectrum estimate.

To preprocess the compressed data, we first compute the sample covariance matrix of the compressed measurements from the available T sample as $\hat{\boldsymbol{R}}_{\boldsymbol{y}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}(t)\boldsymbol{y}^{\mathrm{H}}(t)$. Exploiting the Hermitian property of the covariance matrix, we only consider the upper triangular elements, which are vectorized before feeding into the neural network as $\bar{\boldsymbol{r}} = \mathrm{vec}(\mathrm{triu}(\hat{\boldsymbol{R}}_{\boldsymbol{y}}))$. We then separate and stack the real and imaginary part as $\boldsymbol{r} = [(\mathcal{R}(\boldsymbol{r}))^{\mathrm{T}}\,(\mathcal{I}(\boldsymbol{r}))^{\mathrm{T}}]^{\mathrm{T}}$.

For a total of B observations, the complete training dataset is formed by concatenating the vectors $\boldsymbol{r}_b$ corresponding to the observations $b \in \{1, 2, \cdots, B\}$. The complete training dataset $\boldsymbol{\mathcal{X}}$ is obtained by stacking all B vectors as $\boldsymbol{\mathcal{X}} = [\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_B]$.

The proposed neural network structure is depicted in Fig. 2, where the fully connected network consists of 5 hidden layers. The number of hidden layers and the number of nodes in each layer are experimentally chosen to trade off between the optimization accuracy and the generalization capability of the network. Dropout regularization is introduced at each layer of the network which randomly discards a subset of neural network nodes to reduce overfitting.

Consider the $\mathcal{L}$-layer neural network and use $\boldsymbol{\mathcal{W}}^{[\ell]}$ and $b^{[\ell]}$ to respectively denote the weights and biases of the ℓth hidden layer with $\ell \in \{1, 2, \cdots, \mathcal{L}\}$, and $\mathcal{F}^{[\ell]}$ is the nonlinear activation function applied on the ℓth layer. The output resulted from the ℓth hidden layer of the neural network can be expressed as

$$\boldsymbol{\mathcal{A}}^{[\ell]} = \mathcal{F}^{[\ell]}\left(\boldsymbol{W}^{[l]}\boldsymbol{\mathcal{A}}^{[l-1]} + b^{[l]}\right). \tag{13}$$

As the activation function, we used rectified linear unit (ReLU) in the hidden layer and the sigmoid function in the output layer.

DOA estimation is considered as a multilabel binary classification problem, where the objective is to, for each grid point, make a binary decision whether a signal is present or not. The sigmoid activation function guarantees that the output nodes of the neural network have values between 0 and 1, which can be interpreted as the probability that a signal is present at a particular angle. Therefore, the normalized output of the neural network is regarded as a probability distribution of the signal DOAs and is used as the prior information for the optimization of $\mathbf{\Phi}$ in the subsequent iteration.

A binary cross-entropy loss function is exploited to train a binary classifier. We exploit the Adam (adaptive moment estimation) optimizer [26] to optimize the network parameters in order to minimize the overall cost function, given as

$$\min_{\boldsymbol{W},\boldsymbol{b}} -\frac{1}{J}\sum_{j=1}^{J}\left[\boldsymbol{\mathcal{Y}}_j^{[i]}\log\hat{\boldsymbol{\mathcal{Y}}}_j^{[i]} + \left(1-\boldsymbol{\mathcal{Y}}_j^{[i]}\right)\log\left(1-\hat{\boldsymbol{\mathcal{Y}}}_j^{[i]}\right)\right], \tag{14}$$

where J is the number of training samples in the jth batch, and $\hat{\boldsymbol{\mathcal{Y}}}_j^{[i]}$ and $\boldsymbol{\mathcal{Y}}_j^{[i]}$ are, respectively, the predicted output and the actual label of the ith sample at the jth batch of the training data.

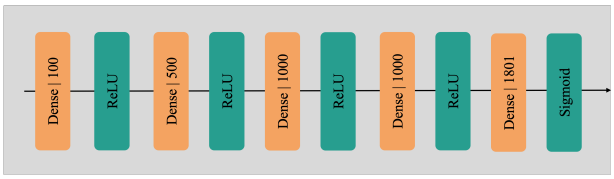


Fig. 2: Proposed neural network structure. The number in each dense layer indicates the number of neurons being used.

5. SIMULATION RESULTS

We consider a massive MIMO system equipped with $N = 50$ antennas which are arranged in a uniform linear fashion and are separated by half-wavelength. 9 uncorrelated far-field signals impinge into the array with input SNR of 20 dB and the number of snapshots is $T = 100$. We choose the compression ratio to be $N/M = 5$, which results in the dimension of the compressed measurement vector to be $M = 10$. The PDF of signal DOA is uniformly discretized with a width of $\Delta\bar{\theta} = 0.1^\circ$, rendering $K = 1801$ components in the Gaussian mixture model. The step size in the iterative optimization of the sensing matrix $\mathbf{\Phi}$ is set as $\gamma = 0.001$.

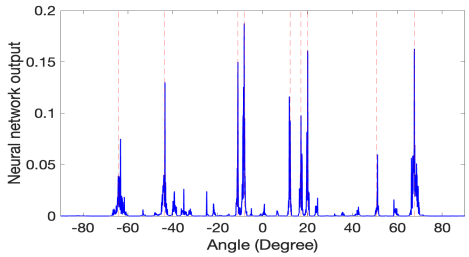
We generate our training dataset by considering the signals impinging from directions within the range of $[-90^\circ, 90^\circ]$. The entire spatial space is discretized with a 0.1° interval, rendering 1801 direction grids. The antenna gains are independently generated from a uniform distribution between 0.5 and 1.5, whereas the phase errors are independently generated from a uniform distribution between -6° and 6°. We randomly take $50,000$ samples of the 9 sources from a uniform distribution to generate the training dataset, and $1,000$ samples are similarly generated as the test dataset.

For $N = 50$ antennas and $T = 100$ snapshots, the dimension of the array received signal is 50×100. Assuming no prior information in the first iteration, we consider the signal DOAs to be uniformly distributed over all angular bins. The compressive measurement matrix $\mathbf{\Phi}$ is optimized based on Eqs. (10) and (11). We then compute the 10×1 compressed measurement vector $\boldsymbol{y}$ using Eq. (3). We then compute the covariance matrices of the compressed measurement vectors for all training data to feed to the neural network. The neural network is trained to learn a nonlinear relationship between the compressed measurements and the signal DOAs by minimizing the binary cross entropy cost function. We use Adam optimizer with a learning rate of 0.001 to train the network. The batch size is set to $M = 64$ and 150 epochs are used for the training.

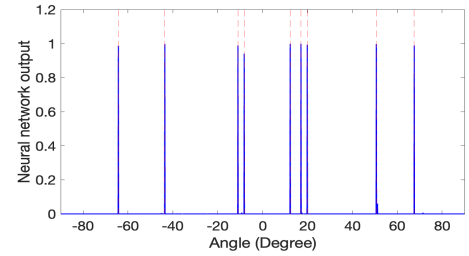
The normalized output of the neural network is regarded as the probability distribution of the signal DOAs and is then used as the prior knowledge for the optimization in the subsequent iteration. Unlike in the first iteration, in the subsequent iterations, we have $50,000$ different prior distributions corresponding to the $50,000$ training observations, each renders a separate optimize matrix $\mathbf{\Phi}$ associated with the specific observation.

In summary, we iteratively optimize the compressive measurement matrix $\mathbf{\Phi}$. In the first iteration, we optimize a generic $\mathbf{\Phi}$ which is applicable for all observations based on a single uniform prior distribution. The neural network provides different outputs for different observations, which are used as prior distributions for the subsequent iteration. As we have specific priors for specific observations in this case, we will obtain a specific $\mathbf{\Phi}$ optimized for that specific observation, and the result is then used to compute the refined compressed measurement vectors.

We used the test dataset to evaluate the effectiveness of the proposed model. As an example, consider a test scenario where signals are impinging on the array with DOAs -64.3°, -43.6°, -11°, -8.2°, 12.3°, 17.1°, 19.9°, 50.7°, and 67.6°. Fig. 3(a) shows the neural network output after the first iteration. Although the result coarsely approximates the DOAs, it does not accurately resolve them due to uniform priors. This coarse spatial distribution is normalized and used as the prior in the next iteration. The results in the second iteration provide an accurate and sharp spatial spectrum, as shown in Fig. 3(b). For comparison, Fig. 4 shows the result obtained using the method presented in [23]. The performance is degraded because this method fails to learn the sensor impairments. It is noted that, in Fig. 3, the Y-axis results represents the neural network output which is obtained form a sigmoid activation function and depicts the probability in each angular grid. As such, it is different to the MVDR spectrum showing in Fig. 4.



(a) Iteration 1



(b) Iteration 2

Fig. 3: Probability of presence of signals in each angular grids.

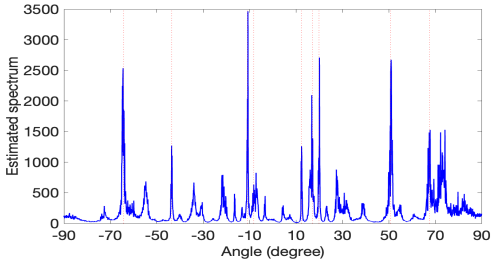


Fig. 4: Performance using the method in [23].

6. CONCLUSION

In this paper, we developed a neural network-based optimization method of the compressive measurement matrix to reduce the number of RF front-end chains in a massive MIMO system. The optimization is based on the maximization of the mutual information between the compressed measurements and the signal DOA without assuming any prior distribution of signal DOAs. The proposed method iteratively updates the signal DOA distribution used for optimization based on the network output of the spatial spectrum. The network is further trained to account for sensor imperfections. The proposed approach provides enhanced performance, faster convergence, and robustness against array imperfections.

7. REFERENCES

[1] R. C. de Lamare, "Massive MIMO systems: Signal processing challenges and research trends," *URSI Radio Science Bulletin*, Dec. 2013.

[2] 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, Jan. 2013

[3] 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, Feb. 2014.

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

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

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

[7] 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, May 2019.

[8] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[9] C. Wang, F. Haider, X. Gao, X. 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, pp. 122–130, Feb. 2014.

[10] 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, Sept. 2017.

[11] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—What is next?: Five promising research directions for antenna arrays," *Digital Signal Process.*, vol. 94, pp. 3-20, 2019.

[12] 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.

[13] S. Sun and Y. D. Zhang, "Multi-frequency sparse array-based massive MIMO radar for autonomous driving," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2020.

[14] V. Cevher, A. Gurbuz, J. McClellan, and R. Chellappa, "Compressive wireless arrays for bearing estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Mar./Apr. 2008, pp. 2497–2500.

[15] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Cardiff, UK, Aug. 2009, pp. 626–629.

[16] A. Gurbuz, J. McClellan, and V. Cevher, "A compressive beamforming method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Mar./Apr. 2008, pp. 2617–2620.

[17] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[18] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," Inverse Prob., vol. 23, no. 3, pp. 969–985, April 2007.

[19] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, March 2008.

[20] 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)*, Vancouver, Canada, May 2013, pp. 6630–6634.

[21] 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)*, New Orleans, LA, March 2017, pp. 3181–3185.

[22] 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, Oct. 2019.

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

[24] S. R. Pavel, W. T. S. 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*, Pacific Grove, CA, Oct. 2021, pp. 241–245.

[25] Z. Liu, C. Zhang, and P. S. Yu, "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, Dec. 2018.

[26] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diago, CA, May 2015, pp. 1–15.