Reinforcement Learning-Based Weak Signal Detection from Compressed Measurements in Massive MIMO Systems

Md. Saidur R. Pavel and Yimin D. Zhang

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

Abstract—**In this paper, we consider optimizing a compressive measurement matrix (CMM) in a massive multiple-input multiple-output (MIMO) system that provides reliable detection capability of both strong and weak signals. To achieve this goal, we propose a reinforcement learning framework, wherein the base station acts as an agent and interacts with the environment to design the CMM by selecting appropriate actions based on a well-defined reward function. Our proposed framework yields improved weak signal detection capabilities. The optimized CMM obtained through the proposed method can then be utilized to reduce the dimension of the received signal, making it practical to implement a massive MIMO system by reducing the number of required radio frequency front-end circuits.**

keywords: Reinforcement learning, weak signal, massive MIMO, compressive measurement matrix, direction-of-arrival estimation.

I. Introduction

Massive multiple-input multiple-output (MIMO) is an important technology for the future-generation wireless communications [1–6]. This technology offers the capability to support a high number of antenna arrays, which can lead to improved system capacity, energy efficiency, and robustness. Additionally, the highly directional beam produced by a massive MIMO system is an effective method of overcoming the propagation delay experienced in millimeter-wave (mmWave) channels [7–9]. Massive MIMO is also receiving increasing interests in radar sensing due to its ability to enhance the sensing capability, coverage, and performance [10, 11].

The implementation of a massive MIMO system necessitates the allocation of radio frequency (RF) front-end circuits and analog-to-digital converters (ADCs) to the antennas for subsequent signal processing. However, the allocation of dedicated RF front-end circuits and an ADC to each antenna is not practical due to a number of factors, including power consumption, hardware complexity, and cost. To address this issue, a hybrid analog-digital processing strategy is effective, allowing for a reduced number of RF circuits and ADCs to be connected to all antennas through a network of phase shifters.

In essence, the analog part of the hybrid beamforming is to reduce the dimension of the received RF signals before they are digitized. In such a scheme, the signals received at the antennas cannot be directly observed, making the optimized design of the analog beamformer difficult. In [12, 13], an information-theoretic approach for the optimization of the compressive measurement matrix is proposed by assuming coarse knowledge of the spatial distributions of the signal arrivals. The requirement of such coarse knowledge is eliminated in [14] through iterative learning. This approach is further extended to a deep learning-based sequential strategy to provide offline-training and generalization capabilities [15]. The optimized compressed measurement matrix yields compressed measurements of the original high-dimensional array signals so the outcomes can be processed with a reduced number of front-end circuits, followed by signal processing methods to obtain interested signal parameters, such as signal directions-of-arrival (DOAs).

The methods described above assume that all signals have an equal or similar strength. As compressed measurements in general favor strong signals, such schemes may likely fail to detect weaker signals. This problem becomes more pronounced in iterative methods, where the probability assigned to weaker signals may become weaker as the iteration continues. There are several studies that address the issue of detecting weak signals in the context of MIMO radar. For instance, [16] improves the detection of weak targets by utilizing the spatial diversity offered by a MIMO radar, whereas [17] enhances weak target detection by utilizing canonical correlation analysis, and [18] proposed a reinforcement learning (RL) approach to improve weak target detection in the context of MIMO cognitive radar.

Driven by the aforementioned considerations, we present an RL framework to improve the detection of weaker signals from compressed measurements in a massive MIMO system. The objective of this paper is to optimize a compressive measurement matrix in which the weak signals are properly preserved. Unlike the iterative approach [14] where the normalized spatial spectrum obtained in an iteration acts as a prior in the subsequent iteration for CMM optimization, the proposed method modifies this prior information before the optimization. After each iteration, angular bins with a lower power spectral density are identified as potential bins containing weak signals. During each iteration, the prior information in the next iteration is adjusted using the State-Action-Reward-State-Action (SARSA) strategy to optimize the CMM, thereby enhancing the system's weak signal detection capability.

Notations: We use lower-case (upper-case) bold characters to describe vectors (matrices). $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ respectively represent the transpose and conjugate transpose of a matrix or vector. $\mathrm{diag}(\cdot)$ denotes a diagonal matrix with the elements of a vector constituting the diagonal entries, whereas $\mathrm{vec}(\cdot)$ denotes vectorizing of a matrix. $\mathrm{triu}(\cdot)$ denotes the upper triangular elements of a matrix. $\mathbb{E}[\cdot]$ denotes the expectation operation. $\circ$

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-23-1-0255. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Air Force.

is the Hadamard product operator. $\text{sign}(x)$ returns 1 if $x \geq 0$ and 0 otherwise. $\jmath = \sqrt{-1}$ denotes the unit imaginary number. In addition, $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ denote the real part and imaginary part of a complex entry, respectively.

II. Signal Model

We consider a massive MIMO system receiver equipped with N receive antennas, and D uncorrelated far-field signals impinge from directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_D]^{\mathrm{T}}$. As shown in Fig. 1, for an array using fully digital beamforming without analog compression, the signal $\boldsymbol{x}^{\mathrm{RF}}(t)$ received at each antenna is processed separately to obtain the baseband signal vector of the array $\boldsymbol{x}(t)$, expressed as

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

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 the 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) vector.

For large-scale antenna arrays in a massive MIMO system, such an approach is impractical due to the high hardware requirements. To address this issue, we project the received signal vector of the array with dimension N onto a lower-dimensional space of dimension M with $M \ll N$. This is achieved by an analog beamformer that associates the output array channels with M measurement kernels, represented by a row vectors $\{\boldsymbol{\phi}_m, \ m = 1, \cdots, M\} \in \mathbb{C}^{1 \times N}$, as shown in Fig. 2. By stacking the measurement kernels, we obtain the compressive sampling matrix $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1^{\mathrm{T}}, \cdots, \boldsymbol{\phi}_M^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{M \times N}$. The matrix entries are converted to analog weights through a digital-to-analog converter (DAC) to perform analog beamforming.

The compressive measurement matrix is used to obtain an M-dimensional compressed measurement vector of the N-dimensional array received signal vector $\boldsymbol{x}(t)$ as

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

The main objective of this paper is to optimize the compressive sampling matrix $\boldsymbol{\Phi}$ with enhanced weak signal detection.

III. Iterative optimization of the compressive measurement matrix $\boldsymbol{\Phi}$

In this section, we review the iterative optimization of the compressive sampling matrix $\boldsymbol{\Phi}$ as discussed in [13, 14]. Consider θ be a random variable that represents the spatial distribution of signal arrivals and is characterized by a probability density function (PDF) denoted as $f(\theta)$. We proceed by discretizing the PDF $f(\theta)$ into K angular bins, each with a width of $\Delta\bar{\theta}$. The yielding probability mass function (PMF) of the kth angular bin is given by $p_k = f(\bar{\theta}_k)\Delta\bar{\theta}$ with $\sum_{k \in \mathcal{K}} p_k = 1$, where $\mathcal{K} = \{1, 2, \cdots, K\}$, and $\bar{\theta}_k$ represents the nominal DOA of the kth angular bin.

The signal arrival from the kth angular bin is modeled with a zero mean Gaussian distribution with variance σ_s^2, i.e., $s(t) \sim \mathcal{CN}(0, \sigma_s^2)$. The compressed measurement at the kth bin with nominal DOA $\bar{\theta}_k$ can then be expressed as

$$y(t)|_{\theta=\bar{\theta}} = \boldsymbol{\Phi}\boldsymbol{a}(\bar{\theta}_k)\boldsymbol{s}(t) + \boldsymbol{n}(t). \tag{3}$$

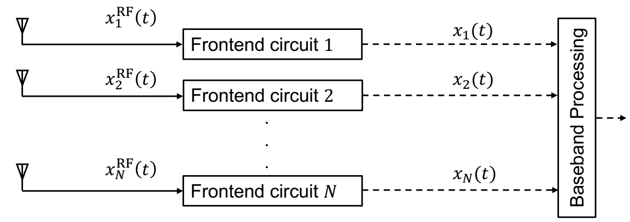


Fig. 1: Block diagram of a massive MIMO receiver without using compression.

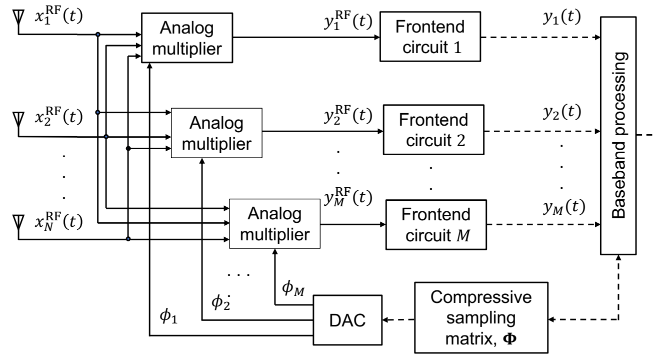


Fig. 2: Block diagram of a massive MIMO receiver exploiting compression for dimension reduction.

In many cellular communication systems and in the tracking mode of radar operations, coarse knowledge of PDF $f(\theta)$ is available. In this case, the PDF of the compressed measurement vector $\boldsymbol{y}(t)$ can be expressed as

$$f(\boldsymbol{y}) = \int f(\boldsymbol{y}|\theta) f(\theta) d\theta \approx \sum_{k \in \mathcal{K}} p_k f(\boldsymbol{y}|\bar{\theta}_k), \tag{4}$$

where the conditional PDF $f(\boldsymbol{y}|\bar{\theta}_k)$ is given as

$$f(\boldsymbol{y}|\bar{\theta}_k) = \frac{1}{\pi^M |\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}|} e^{-\boldsymbol{y}^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}^{-1} \boldsymbol{y}}, \tag{5}$$

with $\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k} = \boldsymbol{\Phi}(\sigma_s^2 \boldsymbol{a}(\bar{\theta}_k) a^{\mathrm{H}}(\bar{\theta}_k) + \sigma_n^2 \boldsymbol{I})\boldsymbol{\Phi}^{\mathrm{H}}$ denoting the covariance matrix of vector $\boldsymbol{y}$ given DOA $\bar{\theta}_k$.

The compressive measurement matrix $\boldsymbol{\Phi}$ is optimized by maximizing the mutual information between the compressed measurement and the signal DOA, denoted as $I(\boldsymbol{y}; \theta)$. This maximization is achieved through a gradient ascent strategy, expressed as

$$\boldsymbol{\Phi} \leftarrow \boldsymbol{\Phi} + \alpha \nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y}; \theta), \tag{6}$$

where $\alpha > 0$ is a learning rate.

When coarse knowledge of $f(\theta)$ is not available, it can be iteratively estimated. Starting with the uniform PMF of θ, denoted as $p^{(0)}(\theta)$, the iterative optimization process optimizes the compressive measurement matrix $\boldsymbol{\Phi}^{(i)}$ at the ith iteration, considering the prior distribution $p^{(i-1)}(\theta)$ for $i \geq 1$. The estimated normalized minimum variance distortionless response (MVDR) spectrum based on the optimized $\boldsymbol{\Phi}^{(i)}$ is then used as the prior distribution of θ in the subsequent iteration. This two-step procedure can be described as

$$P_{\mathrm{MVDR}}^{(i)}(\theta) = \frac{\boldsymbol{a}^{\mathrm{H}}(\theta)(\boldsymbol{\Phi}^i)^{\mathrm{H}}(\boldsymbol{\Phi}^i)\boldsymbol{a}(\theta)}{\boldsymbol{a}^{\mathrm{H}}(\theta)(\boldsymbol{\Phi}^i)^{\mathrm{H}}\left(\hat{\boldsymbol{R}}^{(i)}\right)^{-1}(\boldsymbol{\Phi}^i)\boldsymbol{a}(\theta)} \tag{7}$$

and

$$p_k^{(i+1)} = \frac{P_{\text{MVDR}}^{(i)}(\theta_k)}{\sum_{j\in\mathcal{K}} P_{\text{MVDR}}^{(i)}(\theta_j)}, \quad (8)$$

where $\hat{\boldsymbol{R}}^{(i)}$ is the sample covariance matrix at ith iteration.

In the iterative approach, the PMF of θ converges to have sharp peaks at the true signal DOAs. However, the compressive measurement matrix obtained in such an approach generally favors strong signals. Therefore, when signals have mixed strengths, it may progressively decrease weak signals as the iterations progress and may eventually lead to miss detection. To protect such weak signals, an RL-based framework is considered in the following sections.

IV. Proposed Reinforcement Learning-Based Weak Signal Detection

A. Reinforcement Learning Framework

RL is a machine learning approach that deals with sequential decision-making. It involves an agent that observes its environment and learns to take suitable actions to maximize future rewards. The ultimate goal in RL is to develop an effective control policy or set of actions through positive or negative reinforcement. The folowing key elements are used to characterize the RL process:

- State: A set of observations that describes the environment. The state, denoted as $s^{(i)} \in \mathcal{S}$, represents the observation at iteration i, where $\mathcal{S}$ denotes the set of possible states.
- Action: An action $a^{(i)} \in \mathcal{A}$ represents one of the feasible choices the agent has at iteration i, where $\mathcal{A}$ denotes the set of possible actions. Executing an action results in a transition from the current state, s, to a new state, s'.
- Reward: A reward, denoted as $r^{(i)}$, is achieved by executing action $a^{(i)}$ in a given state $s^{(i)}$. It serves as a performance metric to assess the effectiveness of action $a^{(i)}$ given a state $s^{(i)}$ at itearation i.
- Policy: The policy $\pi(s^{(i)}, a^{(i)})$ represents the probability of selecting action $a^{(i)}$ based on the state $s^{(i)}$.

The agent starts in an initial state of the environment $s^{(0)} \in \mathcal{S}$ based on some observation. At each iteration, it takes an action $a^{(i)} \in \mathcal{A}$ based on a policy function denoted by $\pi(s^{(i)}, a^{(i)})$, which represents the probability of taking action $a^{(i)}$ at state $s^{(i)}$ of the environment. By taking action $a^{(i)}$, two outcomes follows, namely, i) the agent obtaining a reward $r^{(i)} \in \mathcal{R}$, and ii) the state transitions to $s^{(i+1)} \in \mathcal{S}$.

The goal of the RL agent is to find a policy $\pi(s^{(i)}, a^{(i)})$ to maximize an expected return, which is also referred to as the value function, defined as

$$V^{\pi}(s) = \mathbb{E}\left[\sum_{\tilde{i}=0}^{\infty} \gamma^{\tilde{i}} r^{(i+\tilde{i}+1)} | s^{(i)} = s, \pi\right], \quad (9)$$

where $\gamma \in [0,1]$ is called the discount factor, which controls the weights of the future reward. Then, the optimal expected return is

$$V^{\star}(s) = \max_{\pi} \; V^{\pi}(s). \quad (10)$$

In addition to the value function, the Q function associated with Q-learning, a model-free RL method, can be defined in a similar fashion for a particular state and action as

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}\left[\sum_{\tilde{i}=0}^{\infty} \gamma^{\tilde{i}} r^{(i+\tilde{i}+1)} | s^{(i)} = s, a^{(i)} = a\right]. \quad (11)$$

The optimal Q function can be defined as

$$Q^{\star}(s,a) = \max_{\pi} \; Q^{\pi}(s,a), \quad (12)$$

and the optimal policy at state s would be taking an action a that maximizes the $Q(s,a)$, i.e.,

$$\pi^{\star}(s) = \arg\max_{a\in\mathcal{A}} Q(s,a). \quad (13)$$

In other words, the optimal action at a particular state should be the one that maximizes the Q function.

B. Proposed Weak Signal Detection Scheme

To enhance the detection of weak signals, we adopt the SARSA strategy to update the Q function. SARSA is an on-policy RL algorithm that evaluates and improves the same policy used for action selection. It is also a model-free RL algorithm that does not require a model of the environment.

In this particular problem, the objective is to design the CMM $\boldsymbol{\Phi}$. The process for designing $\boldsymbol{\Phi}$ is described below:

1) Selection of state: The CMM $\boldsymbol{\Phi}$ is designed using the mutual information maximization criterion, as described in [13]. When no prior information is available for the signal DOAs, the spatial distribution of the signals is initialized using a uniform distribution [14]. Subsequently, the MVDR spectrum $p^{(i)}(\theta)$ is estimated by eq. (7), exploiting the obtained compressed measurements. The normalized estimated spectrum is treated as the posterior distribution of the DOAs and can be considered an observation of the environment. To obtain the current state, we empirically choose a threshold δ such that

$$\tilde{p}^{(i)}(\theta) = \begin{cases} 1, & \text{if } p^{(i)}(\theta) \geq \delta, \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

$\tilde{p}^{(i)}(\theta)$ signifies candidate angular bins that likely contain signals. The number of angular bins where the spatial spectrum is above the threshold constitutes the state of the environment. As such, the state at the ith iteration denotes the total number of angular bins that likely contain signals and is given as

$$s^{(i)} = \sum_{k=1}^{K} \tilde{p}^{(i)}, \quad (15)$$

with $s^{(i)} \in \{1, 2, \cdots, K\}$.

2) Selection of action: Now, based on the current state at the ith iteration, denoted as $s^{(i)}$, an action $a^{(i)}$ needs to be selected. The action involves two tasks. First, it includes selecting the number of angular bins that most likely contain both strong and weak signals. To accomplish this, we identify an angular bin as possibly containing a weak signal if $\tilde{p}(\theta)$ at that angular bin is less than $\delta/2$. Recognizing these angular bins, the power spectrum is multiplied by a factor of 5, while the power spectrum at the remaining angular bins remains unchanged, i.e.,

$$\hat{p}^{(i)}(\theta) = \begin{cases} p^{(i)}(\theta), & p^{(i)}(\theta) \geq \delta/2, \\ 5p^{(i)}(\theta), & p^{(i)}(\theta) < \delta/2. \end{cases} \quad (16)$$

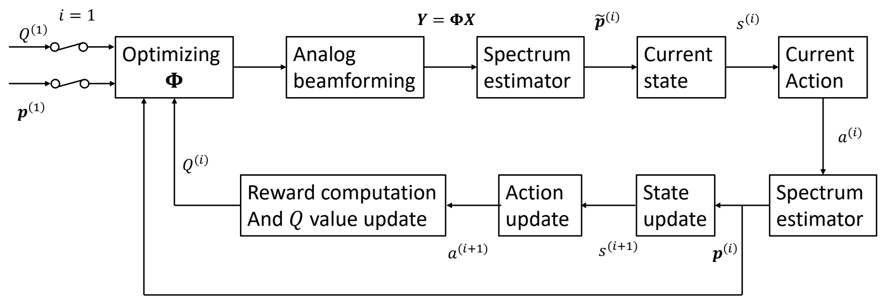


Fig. 3: RL framework for weak signal protection.

By applying a threshold similar to that in equation (14) and calculating the number of angular bins above the threshold, the candidate action is first obtained as

$$a^{(i)}_{\text{candidate}} = \sum_{k=1}^{K} \text{sign}(\hat{p}^{(i)}(\theta) - \delta) . \quad (17)$$

Clearly, $a_{\text{candidate}} \in \{1, 2, \cdots, K\}$ provides a number of angular bins, among which many may not contain any signals. Subsequently, we construct an action space, which is the set of numbers ranging from the state $s^{(i)}$ to the candidate action $a^{(i)}_{\text{candidate}}$. The action space $\mathcal{A}^{(i)}$ in the ith iteration is given as

$$\mathcal{A}^{(i)} = \{s^{(i)}, \cdots, a^{(i)}_{\text{candidate}}\}. \quad (18)$$

The action $a^{(i)}$ involves choosing a number from the action space $\mathcal{A}^{(i)}$. To accomplish this, a state action matrix $\boldsymbol{Q} \in \mathbb{R}^{K \times K}$ with elements $Q(s, a)$ is considered. Each row represents possible states, and each column represents possible actions, i.e., the $(s^{(i)}, a^{(i)})$th element of matrix $\boldsymbol{Q}$ denotes the Q value for taking action $a^{(i)}$ from state $s^{(i)}$. The matrix $\boldsymbol{Q}$ is initialized randomly and updated iteratively. The action $a^{(i)}$ at state $s^{(i)}$ is selected by finding the index for which $\boldsymbol{Q}(s^{(i)}, :)$ has the maximum value.

Based on the particular state at the ith iteration, we employ the ϵ-greedy policy to take an action, i.e,

$$a^{(i)} = \begin{cases} \arg\max\limits_{a \in \mathcal{A}} Q(s^{(i)}, a), & \text{with probability } 1 - \epsilon, \\ \text{random action}, & \text{with probability } \epsilon. \end{cases} \quad (19)$$

Once the action $a^{(i)}$ is chosen, the angular bins corresponding to the highest $a^{(i)}$ values from equation (16) are selected. Let $\Theta^{(i)}_a = \{\theta_1, \theta_2, \ldots, \theta_{a^{(i)}}\}$ with $\hat{p}^{(i)}(\theta_1) \geq \hat{p}^{(i)}(\theta_2) \geq \hat{p}^{(i)}(\theta_{a^{(i)}})$ denoting the power spectra of the selected angular bins. The CMM is then re-optimized by modifying the PMF so that only the angular bins belonging to $\Theta^{(i)}_a$ have nonzero PMFs. Similarly, the set $\Theta^{(i)}_s$ is defined as comprising angular bins with the highest $s^{(i)}$ values from (7) and the set Θ contains all the available angular bins.

3) Computing reward and updating the $\boldsymbol{Q}$ matrix: Exploiting the re-optimized $\boldsymbol{\Phi}$, the MVDR spectrum $p^{(i+1)}$ is subsequently estimated. Similar to the preceding two subsections, the next state $s^{(i+1)}$ and action $a^{(i+1)}$ are chosen based on the $p^{(i+1)}$.

In the SARSA strategy, the updating of $\boldsymbol{Q}$ matrix is described as

$$Q(s^{(i)}, a^{(i)}) \leftarrow Q(s^{(i)}, a^{(i)}) + \alpha(r^{(i+1)} + \gamma Q(s^{(i+1)}, a^{(i+1)}) - Q(s^{(i)}, a^{(i)}), \quad (20)$$

where $\alpha > 0$ is learning rate denoting the extent to which the recent observation overrides the old one, and r is the reward. The reward at iteration i is formulated as

$$r^{(i+1)} = \sum_{\theta_m \in \Theta^{(i)}_s} p^{(i+1)}(\theta_m) - \sum_{\theta_n \in \Theta \setminus \Theta^{(i)}_s} p^{(i+1)}(\theta_n) - |a^{(i)} - s^{(i+1)}|. \quad (21)$$

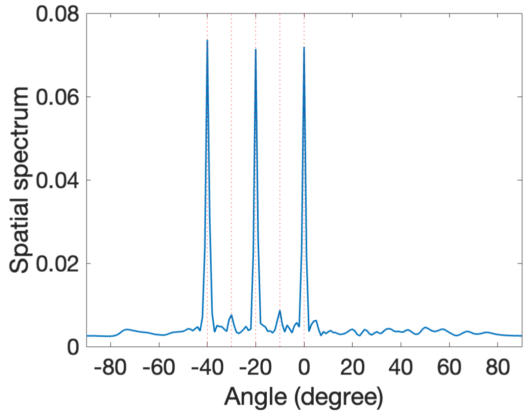
The reward function $r^{(i+1)}$ consists of three components. The first part provides a positive reward for the accurate detection of angular bins from the previous state, which may initially contain the bins corresponding to strong signals. The second term penalizes detection at other angular bins. The third part imposes a penalty if the action from the previous step differs from the current state. Initially, the action denotes the number of angular bins containing strong signals, possibly weak signals, and some bins without containing signals. Since the current state depends on the PMF selection based on these bins, this penalty forces the removal of signal-free bins from the action. If any angular bin containing a weak signal is present in the action, a high PMF assigned to this angular bin may cause the power spectrum at this bin to exceed the threshold, resulting in its detection as a state for the next iteration. If a bin containing a weak signal is detected as a state in the next iteration, the first part of the reward function will ensure the presence of this angular bin in subsequent iterations.

V. Simulation Results

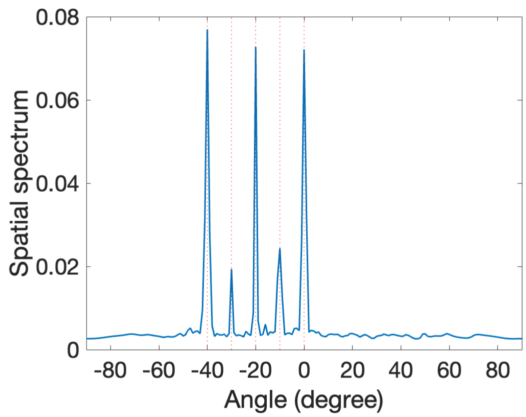
We consider a massive MIMO system consisting of $N = 50$ receive antennas. We choose the compression ratio $N/M = 5$ yielding the dimension of the compressed measurement $M = 10$. The angular bins are discretized with a width of $\Delta\theta = 0.1$.

We consider 5 uncorrelated signals that impinge on the massive MIMO system with DOAs of $-40°$, $-30°$, $-20°$, $-10°$, $0°$. The corresponding signal-to-noise ratios (SNRs) for the 5 signals are considered as 0 dB, -5 dB, 0 dB, -5 dB, and 0 dB, respectively, indicating that two of the signals are weaker than the others.

Fig. 4(a) depicts the estimated spatial spectrum using the method described in [14]. This method iteratively updates the posterior distribution of the DOAs to optimize the CMM. The magnitude of the estimated spectrum in the directions of two weak signals is much lower than their actual level. As such, this method fails to detect weak signals properly. On the other hand, Fig. 4(b) depicts the spatial spectrum based on the proposed RL approach. In this plot, the magnitude of the estimated spectrum in the directions of two weak signals is proportional to the actual signal levels, thereby preserving



(a) Estimated based on [14]



(b) Estimated based on the proposed approach

Fig. 4: Comparison of estimated spatial spectra indicating the weak signal detection performance.

weak signals in the compressive measurement to ensure their successfully detection.

VI. Conclusion

The focus of our paper is to improve the detection of weak signals in a massive MIMO system by leveraging compressed measurements. To achieve this goal, we introduce a reinforcement learning framework that employs a well-designed reward function to enhance the detection process. The simulation results confirm the effectiveness of our proposed method.

VII. 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] 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, 2013.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] S. R. Pavel and Y. D. Zhang, "Deep learning-based compressive sampling pptimization in massive MIMO systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Rhodes Island, Greece, June 2023, pp. 1–5.

[16] X.-Z. Dai, J. Xu, Y.-N. Peng, and X.-G. Xia, "A new method of improving the weak target detection performance based on the MIMO radar," in *Proc. CIE Int. Conf. Radar*, Guangzhou, China, Oct. 2006, pp. 1–4.

[17] M. S. Ibrahim and N. D. Sidiropoulos, "Weak target detection in mimo radar via beamspace canonical correlation," in *Proc. IEEE Sensor Array and Multich. Signal Process. (SAM) Workshop*, Hangzhou, China, June 2020, pp. 1–5.

[18] W. Zhai, X. Wang, M. S. Greco, and F. Gini, "Weak target detection in massive mimo radar via an improved reinforcement learning approach," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Singapore, May 2022, pp. 4993–4997.