Learning Based Reflection Optimization for Practical IRS-Assisted Communications

Mubasher Ahmed Khan, Hamza Ahmed Qureshi, and Yun Hee Kim
Dept. of Electronics and Information Convergence Engineering, Kyung Hee University
{mubasher, hamza.ahmed, yheekim}@khu.ac.kr

Abstract

This paper investigates the performance of an intelligent reflecting surface (IRS)-aided multiple input single output communication system by considering practical hardware limitations in IRS reflection. A deep reinforcement learning (DRL) based algorithm is implemented to maximize the received signal-to-noise ratio (SNR) when the amplitudes of IRS elements depend on their phases.

Ⅰ. Introduction

Intelligent reflecting surface (IRS) has been shown to be a key technology enabling beyond 5G communication networks by providing improved coverage while keeping the power consumption low [1]. To realize the gains promised with an IRS, the reflective coefficients of the IRS elements should be be optimized to match the base station (BS) and user channels. Recently, deep reinforcement learning (DRL) has applied to optimize the IRS reflection [2]. A practical IRS control circuit also undergoes phase-dependent amplitude distortion that has an effect on the performance of the IRS [3]. In this paper, we apply the deep deterministic policy gradient (DDPG) algorithm [4] to optimize the practical IRS phase shifts undergoing reflection impairment to improve the performance.

Ⅱ. System Model and Problem Formulation

A single-user multiple-input single-output (MISO) downlink system as shown in Fig. 1 is considered. The BS consists of M antennas, the IRS is composed of $N = N_x \times N_y$ reflecting elements whilst where N_x and N_y are the number of elements in each row and column of the IRS respectively. The channels from the BS-IRS and IRS-user are denoted as $\mathrm{G} \in \mathbb{C}^{N \times M}$ and $\mathrm{h}_r \in \mathbb{C}^{N \times 1}$ respectively. We assume that the channel between the BS and user is blocked and thus there is no direct BS-user link.

For this system, the signal received at the user can be written as

$$y = \mathrm{h}_r^H \Theta \mathrm{G} \mathrm{w} s + n, \tag{1}$$

where $\Theta = \mathrm{diag}(\theta_1, \theta_2, \cdots, \theta_N)$ is the phase shift matrix at the IRS, $\mathrm{w} \in \mathbb{C}^{M \times 1}$ is the beamforming vector at the BS with the constraint $\|\mathrm{w}\|^2 \le P_{\max}$, $P_{\max}$ is the maximum transmit power of the BS, s is the transmitted signal, and $n \sim CN(0, \sigma^2)$ is the noise. For a practical IRS control circuit, the IRS reflection can be expressed as

$$\theta_n = \beta_n(\phi_n) e^{j\phi_n}, \tag{2}$$

where

$$\beta_n(\phi_n) = (1 - \beta_{\min}) \left(\frac{\sin(\phi_n - \phi_0) + 1}{2} \right)^{\alpha} + \beta_{\min,} \tag{3}$$

with $\phi_n \in [0, 2\pi]$, $\alpha \ge 0$, $\beta_{\min} \ge 0$, and $\phi_0 \ge 0$. The received signal-to-noise ratio (SNR) for this system can then be obtained as

$$\gamma(\Theta, \mathrm{w}) = \left| \mathrm{h}_r^H \Theta \mathrm{G} \mathrm{w} \right|^2 / \sigma^2. \tag{4}$$

The most effective beamforming method to maximize the received SNR for a given phase shift matrix Θ is the maximum-ratio transmission (MRT); The beamforming vector is given by

$$\mathrm{w}^*(\Theta) = \sqrt{P_{\max}} \frac{(\mathrm{h}_r^H \Theta \mathrm{G})^H}{\| \mathrm{h}_r^H \Theta \mathrm{G} \|} \tag{5}$$

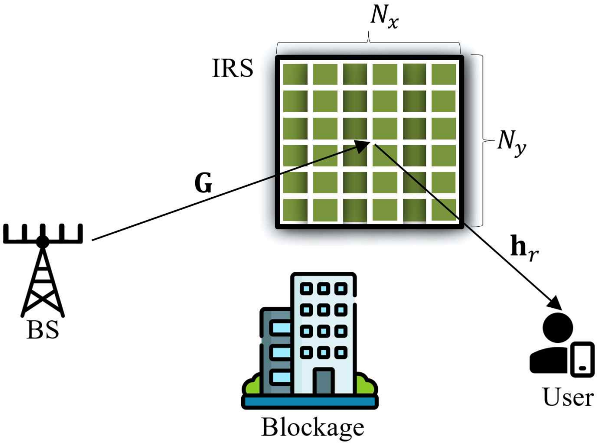

Fig. 1. IRS-assisted MISO system.

and the corresponding SNR is given by

$$\gamma(\Theta, \mathrm{w}^*(\Theta)) = P_{\max} \| \mathrm{h}_r^H \Theta \mathrm{G} \|^2 / \sigma^2. \tag{6}$$

Thus, the IRS reflection optimization is equivalent to

$$\max_{\Theta} \quad \| \mathrm{h}_r^H \Theta \mathrm{G} \|^2 \tag{7a}$$

$$\text{s.t.} \quad 0 \le \phi_n \le 2\pi \tag{7b}$$

This problem is an NP-hard problem as the objective function is a non-convex function. Considering the practical reflection scenario increases the complexity of the problem. We propose employing a DRL based framework to efficiently address this problem.

Ⅲ. DRL Based Framework

In a reinforcement learning system, there are two primary components: the agent and the environment. For each time step t, the agent takes as input the current state information s_t and outputs an action a_t. This action is used to calculate the reward r_t and the state is updated to get the next state s_{t+1}. The algorithm executes for N episodes, and within each episode, it undergoes T iterations or steps. Our optimization problem requires us to generate actions from a continuous space, therefore we employ the DDPG algorithm, which is a model-free, off-policy actor-critic algorithm, as it has been shown to provide good performance in environments with a continuous action space. It employs an actor-critic structure with two neural networks, where the actor suggests actions and the critic evaluates them. Through experience replay and target networks, DDPG stabilizes training, and by outputting deterministic actions, it navigates continuous action spaces. The algorithm updates the policy using policy gradient methods and Q-value updates, striking a balance between exploration and exploitation for effective learning in complex environments.

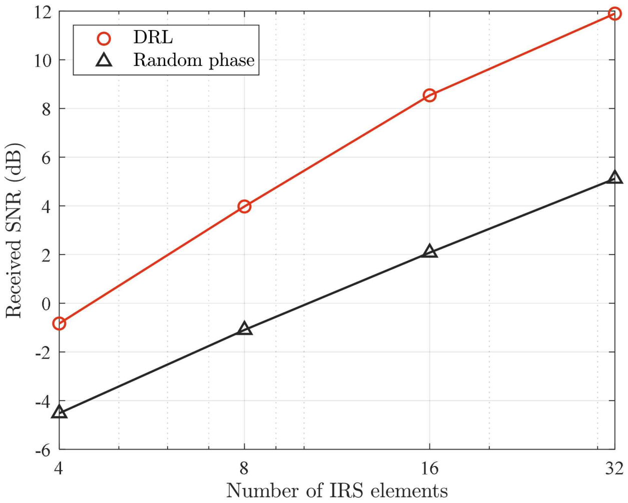

Fig. 2. Received SNR vs BS-user horizontal distance

To be able to use DDPG for our optimization problem, we need to define our state space, action space and reward. For the given system, the communication system can be regarded as the environment and the IRS can be considered as the agent. The construction of the state space, action space and the reward is as follows:

1. *State space*: The state space provides a description of the environment at the current time step. We define the state s_t as
$$s_t = \left[\gamma^{(t-1)}, \phi_1^{(t-1)}, \phi_2^{(t-1)}, \cdots, \phi_N^{(t-1)}\right], \tag{8}$$
where $\gamma^{(t-1)}$ is the received SNR at time step $t-1$.
2. *Action space*: The agent uses the state s_t at each time step t to output the new phase shifts for the IRS elements. The action is therefore defined as
$$a_t = \left[\phi_1^{(t)}, \phi_2^{(t)}, \cdots, \phi_N^{(t)}\right]. \tag{9}$$
3. *Reward function*: The objective of this paper is to maximize the received SNR, thus the reward function is chosen as the received SNR defined in (2). The output of the action network is used to calculate the reward at each time step t.

At the start of each episode, the channel state information is obtained which includes the BS-IRS channel and the IRS-user channel. The first action vector a_0 is initialized with random phase shifts for the IRS to obtain the initial state s_1. This is then used to generate the subsequent actions, rewards and states.

IV. Results and Discussion

For our simulations, we take $M=2$ antennas at the BS, $P_{\max}=15\text{dBm}$, and $\sigma^2=-90\text{dBm}$. The positions of the BS, IRS and user are $(0,0,0)$, $(51,0,0)$, and $(45,1.5,0)$ meters, respectively. To model the practical IRS reflection, we take $\alpha=1.6$, $\beta_{\min}=0.2$, and $\phi_0=0.43\pi$ as in [3]. For the DDPG algorithm, the maximum number of steps per episode were set to 10,000. The learning rate for all neural networks was set to 0.001. The target networks were updated with a decaying rate of 10^{-5}. Buffer size for the experience replay was set to 100,000 and each mini-batch consisted of 16 samples. Simulation results are obtained by averaging over 500 realizations of the random components in the channels.

In Fig. 2, we compare the performance as the number of IRS elements increases. As can be seen from the simulation results, the DRL based agent provides significant gain to the performance of the IRS. Typical convex optimization methods have a high complexity and the computing time for these methods increases exponentially with the increase in number of IRS elements. The complexity of the DRL based framework does not increase significantly with the number of IRS elements and can therefore be scaled to utilize a larger IRS to increase performance.

Acknowledgment

This work was supported by the Ministry of Science and ICT, South Korea, in part through the National Research Foundation of Korea (NRF) under Grant NRF-2021R1A2C1005869 and in part through the Institute for Information & Communications Technology Planning & Evaluation (IITP) under the Information Technology Research Center Support Program under Grant IITP-2023-2021-0-02046.

References

[1] Q. Wu and R. Zhang, “Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network,” *IEEE Commun. Magazine,* vvol. 58, no. 1, pp. 106-112, Jan 2020.

[2] K. Feng, Q. Wang, X. Li and C. -K. Wen, "Deep Reinforcement Learning Based Intelligent Reflecting Surface Optimization for MISO Communication Systems," in *IEEE Wirel. Commun. Letters,* vol. 9, no. 5, pp. 745-749, May 2020.

[3] S. Abeywickrama, R. Zhang, Q. Wu and C. Yuen "Intelligent Reflecting Surface: Practical Phase Shift Model and Beamforming Optimization," in *IEEE Trans. Commun.,* vol. 68, no. 9, pp. 5849-5863, Sept. 2020.

[4] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. 4th Int. Conf. Learn. Represent. (ICLR),* San Juan, PR, USA, May 2016, pp. 1-14.