Kalpa Publications in Computing

Volume 14, 2023, Pages 28–31

Proceedings of V XoveTIC Conference. XoveTIC 2022

kalpa Computing

Joint Optimization of IRS-assisted MIMO Communications through a Deep Contextual Bandit Approach

Dariel Pereira-Ruisánchez, Óscar Fresnedo, Darian Pérez-Adán, and Luis Castedo

Department of Computer Engineering & CITIC Research Center, University of A Coruña, Spain
d.ruisanchez@udc.es, oscar.fresnedo@udc.es, d.adan@udc.es, luis.castedo@udc.es

Abstract

The multiple-input multiple-output (MIMO) communications and the intelligent reflecting surfaces (IRSs) have been envisioned as key technologies for beyond 5G mobile networks. However, the computational complexity of conventional approaches to jointly optimize IRS-assisted MIMO communication systems constitutes a major limitation to their deployment. In this paper, we present an innovative contextual bandit (CB)-based approach for the optimization of the MIMO precoders and the IRS phase-shift matrix entries. The proposed optimization framework, termed as deep contextual bandit-oriented deep deterministic policy gradient (DCB-DDPG), considers a CB formulation with continuous state and action spaces. The simulation results show that our proposal performs remarkably better than state-of-the-art heuristic methods in high-interference scenarios.

1 Introduction

An IRS is a large planar array composed of passive scattering elements having specially designed physical structures which can be individually controlled in a software-defined manner. This allows modifying the phases of the impinging signals to enhance the performance of wireless links. The research interest in IRS-assisted MIMO communication systems is increasing as they stand as an appealing technology to fulfill the requirements of emerging applications [1]. However, the joint optimization of the MIMO precoders and the IRS phase-shift matrix entries is a high-complexity problem. In most scenarios, conventional algorithms are too computationally complex and the search spaces are too vast for considering approaches like the genetic algorithms. Recently, CB is receiving attention and several CB-based solutions have been proposed for open problems in wireless communication systems [2]. This learning-by-interacting approach enables to handle high-dimensionality problems while offering affordable performances in terms of latency and computational complexity. However, all existing works consider discrete action spaces, which is a major limitation for the optimization problem we are dealing with. This issue, and the limitations of the scarce solutions found in the literature, have motivated our work whose main contribution is the development of an actor-critic framework called DCB-DDPG which enables to efficiently handle the continuous action space formulation required to address this optimization problem.

2 System Model

Let us consider the uplink of an IRS-assisted multi-stream (MS) multi-user (MU) MIMO communication system where K users employ N_{t} antennas each to send N_{s} data streams to a base station (BS) with N_{r} antennas with the help of an IRS with N elements. We assume that there is a total blockage between the users and the BS. Hence, the joint optimization of the MIMO precoders and the IRS phase-shift matrix, in terms of sum-rate maximization, is formulated as

$$\begin{aligned} \underset{\mathbf{P},\boldsymbol{\Theta}}{\arg\max} \sum_{k=1}^{K} \log_2 \det \left(\mathbf{I}_K + \mathbf{X}_k^{-1}\mathbf{W}_k^H\mathbf{H}_{\mathrm{IB}}\boldsymbol{\Theta}\mathbf{H}_{\mathrm{UI}k}\mathbf{P}_k\mathbf{P}_k^H\mathbf{H}_{\mathrm{UI}k}^H\boldsymbol{\Theta}^H\mathbf{H}_{\mathrm{IB}}^H\mathbf{W}_k\right) \\ \text{s.t. } \| \mathbf{P}_k \|_F^2 \leq \Omega_k, \ \forall k, \text{ and } \boldsymbol{\Theta} \in \mathcal{D}, \end{aligned} \tag{1}$$

where

$$\mathbf{X}_k = \sum_{i \neq k} \mathbf{W}_k^H\mathbf{H}_{\mathrm{IB}}\boldsymbol{\Theta}\mathbf{H}_{\mathrm{UI}i}\mathbf{P}_i\mathbf{P}_i^H\mathbf{H}_{\mathrm{UI}i}^H\boldsymbol{\Theta}^H\mathbf{H}_{\mathrm{IB}}^H\mathbf{W}_k + \sigma_{\mathrm{n}}^2\mathbf{W}_k^H\mathbf{W}_k \tag{2}$$

is the interference plus noise matrix, and

$$\mathbf{W}_k^H = \mathbf{P}_k^H\mathbf{H}_{\mathrm{UI}k}^H\boldsymbol{\Theta}^H\mathbf{H}_{\mathrm{IB}}^H\left(\mathbf{H}_{\mathrm{IB}}\boldsymbol{\Theta}\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H\boldsymbol{\Theta}^H\mathbf{H}_{\mathrm{IB}}^H + \sigma_{\mathrm{n}}^2\mathbf{I}_{N_{\mathrm{r}}}\right)^{-1} \tag{3}$$

is the minimum mean square error (MMSE) individual receiving filter for the k-th user. $\mathbf{H}_{\mathrm{IB}} \in \mathbb{C}^{N_{\mathrm{r}} \times N}$ and $\mathbf{H}_{\mathrm{UI}k} \in \mathbb{C}^{N \times N_{\mathrm{t}}}$ stand for the channel responses from the IRS to the BS and from the k-th user to the IRS, respectively. $\mathbf{P}_k \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{s}}}$ stands for the k-th user precoder which, as stated in (1), has a power constraint (Ω_k). $\mathbf{H}_{\mathrm{UI}k}, \forall k$ and $\mathbf{P}_k, \forall k$ can be written in a compact manner as $\mathbf{H} = [\mathbf{H}_{\mathrm{UI}1}, \ldots, \mathbf{H}_{\mathrm{UI}K}]$ and $\mathbf{P} = \mathrm{blkdiag}(\mathbf{P}_1, \ldots, \mathbf{P}_K)$, respectively. The IRS phase-shift matrix is represented by the diagonal matrix $\boldsymbol{\Theta} = \mathrm{diag}(e^{j\theta_1}, \ldots, e^{j\theta_N}) \in \mathcal{D}$ with $\theta_n \in [0, 2\pi)$, and $\mathcal{D} \in \mathbb{C}^{N \times N}$ is the set of diagonal matrices with unit modulus entries. The vector that contains the entries in the main diagonal of $\boldsymbol{\Theta}$ is denoted by $\boldsymbol{\theta} \in \mathbb{C}^{N \times 1}$.

3 Deep Contextual Bandit-based Joint Optimization

CB problems can be interpreted as a relaxation of reinforcement learning (RL) problems where interactions can be defined through a dynamics function $p(r_t|a_t, s_t)$, i.e., only immediate rewards are affected by the current state and action. According to the CB formulation, we define that:

- The **state** vector $\mathbf{s}_t$ is composed of the current values of the channel response matrices ($\mathbf{H}_{\mathrm{UI}k}, \forall k$ and $\mathbf{H}_{\mathrm{IB}}$), such that $\mathbf{s}_t = [\mathrm{vec}(\mathbf{H}_{\mathrm{UI}1}), \ldots, \mathrm{vec}(\mathbf{H}_{\mathrm{UI}K}), \mathrm{vec}(\mathbf{H}_{\mathrm{IB}})]$. The vect$(\cdot)$ operator reshapes its input into a row vector.
- The **action** vector $\mathbf{a}_t$ is composed of the entries in the main diagonal of the IRS phase-shift matrix ($\boldsymbol{\theta}$) and those in all the user individual precoders ($\mathbf{P}_k, \ \forall k$). The action vector is hence constructed such that $\mathbf{a}_t = [\mathrm{vec}(\mathbf{P}_1), \ldots, \mathrm{vec}(\mathbf{P}_K), \mathrm{vec}(\boldsymbol{\theta})]$.
- The **reward** r_t is the system sum-rate since this is the metric we aim to maximize.

CB-based algorithms are mostly oriented to solve optimization problems with discrete action formulations. Hence, to account for continuous actions, we propose an innovative CB framework termed as DCB-DDPG since it is inspired on the deep reinforcement learning (DRL)-based DDPG approach. We have considered an actor-critic agent where the critic is trained to learn the immediate reward function, while the actor is trained through deterministic policy gradient updates to predict the action that maximizes the reward function in a given state. Artificial neural networks (ANNs) were employed for both function approximations.

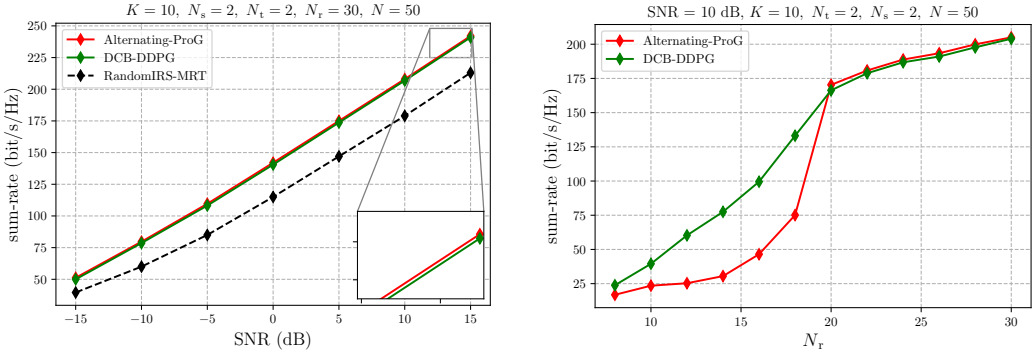

Figure 1: Sum-rate vs signal-to-noise ratio (SNR) (left) and sum-rate vs N_r (right).

4 Simulation Results

We have considered two model-driven approaches to be used as benchmarks during the evaluation of the proposed framework. The first scheme is termed as Alternating-ProG and uses an alternating minimization projected gradient-based algorithm [3]. In the second, termed as RandomIRS-MRT, the IRS phase-shift matrix is randomly selected from $\mathcal{D}$ and the precoders are designed according to the maximum ratio transmitter (MRT) approach.

Figure 1 (left) shows the achievable sum-rates obtained with the proposed DCB-DDPG approach in a scenario where $N_r >> KN_s$. The performance of our proposal is close to the Alternating-ProG approach, whose performance is near optimal in this kind of setups. Besides, our proposal significantly outperforms the RandomIRS-MRT baseline strategy. On the other hand, Figure 1 (right) shows the impact on the system performance of the relationship between the number of receiving antennas N_r and the number of transmitted streams KN_s. As can be seen in the figure, the performance of the Alternating-ProG approach remarkably degrades when N_r decreases below KN_s. When moving into this regime, the Alternating-ProG strategy is not able to properly handle the interference among the users. However, our DCB-DDPG proposal manages the interference more efficiently and the sum-rate values decrease more slowly when reducing the number of receiving antennas at the BS.

5 Conclusions

We have developed a CB-based framework termed as DCB-DDPG to handle the joint optimization of the IRS phase-shift matrix and the user precoders in the uplink of IRS-assisted MS MU-MIMO communications. The simulation results show that this optimization problem can be properly formulated as a CB problem with continuous state and action spaces. Besides, the proposed DCB-DDPG framework stands as an effective method to manage this formulation. It achieves a performance close to the one obtained with state-of-the-art heuristic algorithms when $N_r \geq KN_s$ whereas it handles the MU interference more efficiently when $N_r < KN_s$.

References

[1] Shimin Gong, Xiao Lu, Dinh Thai Hoang, Dusit Niyato, Lei Shu, and Ying-Chang Liang. Toward Smart Wireless Communications via Intelligent Reflecting Surfaces: A Contemporary Survey. *IEEE Communications Surveys and Tutorials*, 22(4):32, 2020.

[2] Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: an introduction*. Adaptive computation and machine learning series. The MIT Press, Cambridge, Massachusetts, second edition, 2018.

[3] Darian Pérez-Adán, Óscar Fresnedo, José P. González-Coma, and Luis Castedo. Alternating Minimization Algorithm for Multiuser RIS-assisted MIMO Systems. In *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, June 2022. ISSN: 2155-5052.