Optimal UAV-Trajectory Design in a Dynamic Environment Using NOMA and Deep Reinforcement Learning

Fatemeh Banaeizadeh, Michel Barbeau, Joaquin Garcia-Alfaro, Evangelos Kranakis

Optimal UAV-Trajectory Design in a Dynamic Environment Using NOMA and Deep Reinforcement Learning

Fatemeh Banaeizadeh*, Michel Barbeau*, Joaquin Garcia-Alfaro†, Evangelos Kranakis*

* School of Computer Science, Carleton University, Ottawa, Ontario, Canada, K1S 5B6
Email: FatemehBanaeizadeh@cmail.carleton.ca, {barbeau,kranakis}@scs.carleton.ca

† Institut Polytechnique de Paris, Telecom SudParis, 91120 Palaiseau, France
Email: joaquin.garcia_alfaro@telecom-sudparis.eu

***Abstract*—Effective deployment of cellular-connected UAV networks necessitates efficient techniques to minimize mutual interference between UAVs and ground users. Moreover, the existing sub-6 GHz band suffers from extreme congestion, making it challenging to allocate unused resource blocks (RBs) for UAVs. This paper presents a learning-based UAV-path planning approach at the Base Station (BS) side, leveraging Non-Orthogonal Multiple Access (NOMA) and Deep Q-Network (DQN) methodologies to address massive connectivity and air-to-ground interference. The proposed NOMA-DQN learning approach optimizes UAV-transmission power and RB allocation jointly, taking into account the UAV-location. Additionally, it devises an interference-aware path for the UAV, considering its limited battery capacity. Simulation results demonstrate the efficacy of our proposed approach in terms of maximizing the total sum rate of aerial and ground users in a shared RB, as well as enhancing UAV energy efficiency, as compared to shortest path, orthogonal multiple-access (OMA), and random selection schemes.**

***Index Terms*—Cellular-connected UAVs, deep reinforcement learning, NOMA, effective energy-consumption, interference-aware trajectory.**

I. Introduction

Uncrewed Autonomous Vehicles (UAVs) are expected to play a significant role in the development of 5G wireless communications and cellular networks [1]. Their ability to accomplish diverse missions such as surveillance, cargo delivery, geographical mapping, and disaster management has dramatically increased demands for the use of UAVs in various domains. UAVs in cellular networks can operate as flying Base Stations (BSs), relay nodes, or aerial users. Using them as flying Base Stations (BSs) enhances coverage, throughput, and reliability of uplink or downlink communications for terrestrial users. Also, a UAV as an aerial user eases information collection from target areas [1], [2]. Apart from their benefits, UAV-effective deployment entails addressing technical challenges.

Since UAVs fly in high altitudes, Line-of-Sight (LoS) channel communication between UAVs and ground BSs, in cellular-connected UAV networks, causes severe interference on their co-channel ground users, exacerbating performance of their uplink communication [3]. This challenge cannot be simply resolved by assigning unused Resource Blocks (RBs) to the UAVs. The number of available RBs in 5G cellular networks is limited [4]. Therefore, effective air-to-ground interference management techniques should be designed. Besides, UAVs have limited battery resources. Maximization of UAV-energy-efficiency is also critical to prolong flight time and accomplish missions. Hence, the design of interference and energy-aware 3D trajectories is of high significance. It is the main focus of this study.

A. Related Work

Different approaches have been proposed to address the problem. The capability of Non-Orthogonal Multiple Access (NOMA) in UAV-uplink interference cancellation on its co-channel ground users is evaluated in [5], [6], considering static UAV-locations. However, the mobility of aerial and ground users is not investigated.

Utilizing directional antennas presents another effective solution to mitigate the adverse effects of UAV-interference by directing signals away from unwanted areas. Equipping the UAV with a directional antenna enables control over aerial interference [7]. Resource optimization is of importance in interference cancellation and authors in [8] address inter-drone interference and aerial-ground interference using scheduling scheme and transmission power optimization in a shared spectrum, respectively. Nevertheless, the effectiveness of the proposed methods for UAV-energy consumption remains an open question. An inverse Reinforcement Learning (RL) approach is proposed in [9] to address aerial interference through joint optimization of power allocation and trajectory, although massive connectivity is not investigated. A local interference cancellation (IC) technique is proposed in [10]. Ground BSs tackle air-to-ground interference, taking into account the static locations of ground users and a UAV flying at a fixed altitude. They manage the interference by joint optimization of users' transmission power and UAV-trajectory design. They disregard the mobility of ground users, also flying UAVs at different altitudes. In our previous work [11], we proposed a learning framework using Multi-armed Bandit (MAB) and NOMA to mitigate the effects of UAV-interference on co-channel ground

users. The BS as an agent employs the MAB-NOMA learning approach to find the best RB and transmission power for every UAV-position to be paired with a terrestrial user. However, mobility of ground users, UAV-path planning and energy consumption are not investigated.

B. Contributions

The high-dynamic environment of cellular-connected UAV networks demands learning approaches capable of fast adaptations. Deep Reinforcement Learning (DRL) offers this feature through continuous environmental interaction [12]. We propose a learning-based NOMA-Deep Q-Network (DQN) framework for UAV trajectory design, addressing the following challenges. The framework minimizes UAV-uplink interference on co-channel ground users by jointly optimizing UAV-transmission power and RB allocation in a dynamic environment, i.e., ground users follow a random walk mobility model. Based on ground user and UAV locations, at each time step, the agent (BS) applies the NOMA-DQN framework to acquire the environment dynamic and choose the best action (UAV-transmission power and RB, paired with a ground user, while moving from an initial point to a terminal point). The optimal action leads to minimizing the UAV-interference and maximizing the sum of data rates of aerial and ground users in a shared RB. Moreover, the framework addresses massive connectivity in dense networks with many ground users, which is crucial for achieving low-latency. That is, if there is no free orthogonal RB to be allocated to a UAV, an agent can dynamically pair the UAV with a ground user using NOMA, ensuring that their minimum requirements are fulfilled. Besides, the framework optimizes the energy-consumption of the UAV by designing a path that minimizes the distance it travels. The proposed optimization function gives rise to a trade-off between maximizing the sum of data rates of aerial and ground users, sharing a RB, and maximizing energy-efficiency of the UAV. The simulation results show that the framework deals with dynamic environments and finds optimal 3D interference minimization and energy-aware paths for a UAV.

The rest of this paper is organized as follows. Section II illustrates our proposed system and communication model, and provides the problem formulation using a DQN architecture. Section III presents the simulation results, followed by a conclusion in Section IV.

II. Proposed Approach

We first describe the system and communication models. Then, the UAV-interference minimization problem is formulated using the NOMA-DQN learning framework to obtain energy-efficient and interference-aware trajectories for a UAV. Finally, the architecture of DQN and the used parameters are explained.

A. System and Communication Models

We consider a single-cell cellular-connected UAV network. It is composed of *n* ground users, a rotary-wing UAV acts as an aerial user, and a BS. The BS is located at the cell center. It acts as a learning agent and serves aerial and ground users. Orthogonal RBs are assigned to terrestrial users. For simplicity, the entire 3D area is represented as a rectangle with dimensions $G_x, G_y, G_z \in \mathbb{N}^+$. It is divided into equal size of grid cells.

Mobility of ground users is modeled according to random walk [13]. That is, a ground user starts form an initial random location. At each instant t, a new direction θ_t is randomly chosen in the interval $(0, 2\pi]$ and a speed v_t is randomly chosen in the interval $[v_{min}, v_{max}]$. The new location becomes $x_t = x_{t-1} + v_t \cos(\theta_t)$ and $y_t = y_{t-1} + v_t \sin(\theta_t)$.

A UAV-flight time *T* is discretized into *N* time steps. It starts from an initial position L_0 at time $t = 0$. It reaches a terminal position L_N at time T, with a set of pre-defined move directions in a 3D space, $\{up, down, east, west, north, south\}$, and a constant speed v_{uav}. The grid position of the UAV at time *t* is denoted by (x_t, y_t, z_t), where z_t is the altitude. At any time *t*, the agent chooses an action a_t that consists of three elements $\{a_t^1, a_t^2, a_t^3\}$. Where a_t^1 is representing a move direction, a_t^2 a transmission power, and a_t^3 a RB for the UAV, to be paired with a ground user. The new location of the UAV is updated according to the action taken by the BS-agent. For example, if the action is *up*, then the new position is equal to $x_t = x_{t-1}$, $y_t = y_{t-1}$, and $z_t = z_{t-1} + v_{uav}t$. The altitude of the UAV is must be between h_{min} and h_{max}.

For a terrestrial single-antenna user *j* at grid position *i*, the uplink channel between the user and BS is affected by large and small-scale fadings. The path loss at location *i* is determined by the equation [14]:

$$PL_{i,j} = PL_0 + 10\alpha \log_{10} d_{i,j} + X_\sigma \quad dB \tag{1}$$

where PL_0 is the path loss at reference distance one meter, α is the path loss exponent, $d_{i,j}$ is the distance between user *j* and the BS, and X_σ denotes a shadowing effect modeled using a Gaussian distribution with zero mean and variance one. As a result, the channel gain of the ground user *j* is equal to:

$$h_{j,i} = g_{j,i} 10^{-PL_{i,j}/10} \tag{2}$$

where $g_{j,i}$ is an independent and identically distributed (iid) exponential random variable with zero mean and variance 1, modeling the small-fading Rayleigh effect.

The UAV is equipped with a single-antenna. The air-to-ground channel between the UAV and BS, for every grid position *i*, is modeled as the following free-space path loss equation [15]:

$$h_{UAV,i} = \frac{\rho_0}{d_{i,UAV}^2} \tag{3}$$

where ρ_0 is the channel power gain at reference distance one meter and $d_{i,UAV}$ is the 3D Euclidean distance (m) between the UAV and BS.

B. Problem Formulation

We apply the NOMA-DQN framework (Algorithm 1) to produce an optimal UAV-flight path. The goal of the produced path is to minimize the energy consumption of the UAV and

to minimize the UAV-uplink interference on the co-channel ground user. The goal is reached finding, step-by-step, the best move direction, transmission power, and RB for the UAV, flying from L_0 and reaching the destination L_N. In standard DRL frameworks [16], an agent solves the problem through sequential decision-making and environmental interactions. At each time *t*, the agent maps a state s_t to an action a_t according to the policy π, that is, $a_t = \pi(s_t)$. In this paper, the state space $\mathcal{S}$ contains the locations of terrestrial and aerial users. Each state $s_t \in \mathcal{S}$ is denoted by:

$$s_t = [(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n), (x_{uav}, y_{uav}, z_{uav})] \quad (4)$$

where $(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_n, y_n, z_n)$ are 3D coordinates of ground users. $(x_{uav}, y_{uav}, z_{uav})$ are the 3D coordinates of the UAV. The action space $\mathcal{A}$ is composed of three discrete sub-action spaces defined as:

$$\begin{aligned} \mathcal{A}_1 &= \{up, down, east, west, north, south\} \\ \mathcal{A}_2 &= \{p_1, p_2, \ldots, p_{max}\} \\ \mathcal{A}_3 &= \{RB_1, RB_2, \ldots, RB_n\} \end{aligned} \quad (5)$$

where $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ represent UAV-move directions, transmission power values, and RBs. Hence, the set of all available action pairs is equal to $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3$. An action $a_t \in \mathcal{A}$, taken by the agent at time t, is defined as $a_t = \{a_t^1, a_t^2, a_t^3\}$.

The reward r_t resulting from the taken action is calculated according to three parameters: uplink NOMA system [17], total distance travelled by the UAV at time *t*, and UAV-distance to L_N. For a NOMA-uplink communication, a BS receives a signal combining all users in the shared RB. It employs the Successive Interference Cancellation (SIC) process to decode the user signals following their received strength, from the strongest to the weakest. Let *n* be the number of available RBs, which is also equal to the number of ground users. Sorting channel gains from highest to lowest in the order $h_1, h_2, \ldots, h_n$, with transmit power levels $p_1, p_2, \ldots, p_n$, the Signal-to-Interference-Noise Ratio (SINR) of user j in the shared RB is computed as:

$$SINR_j = \frac{p_j |h_j|^2}{\sum_{k=j+1}^{n} p_k |h_k|^2 + N_0 (B/n)} \quad (6)$$

where N_0 is noise power density, B/n is the bandwidth (Hz) of the shared RB. The minimum requirement of users in a shared RB is met when their SINRs are greater than or equal to a defined threshold. As a result, the total uplink data rate of all the users in the shared RB is defined as:

$$R_t = \frac{B}{n} \sum_{j=1}^{n} \log_2(1 + SINR_j) \quad \text{bits/second} \quad (7)$$

On the other hand, efficient UAV-energy consumption is achieved by minimizing the total distance travelled by the UAV. Therefore, the agent should produce a path making a trade-off between interference mitigation and energy efficiency. In the proposed framework, the UAV-energy consumption is minimized by reducing the total UAV-travel distance. In other words, the UAV-path from L_0 to L_N is defined as sequential visited locations denoted as $p = (L_0, L_1, L_2, \ldots, L_N)$. The distances between every pair of locations are represented by $(w_1, w_2, \ldots, w_N)$. The total distance from the initial location to the target is equal to $D = \sum_{k=1}^{N} w_k$ meters. The UAV-energy consumption is minimized when the agent finds an optimal path D^* with minimum total distance while interference is also mitigated. It is defined as:

$$D_t = \sum_{k=1}^{t} w_k. \quad (8)$$

Let us define the reward at time t as:

$$r_t = \eta_1 \cdot \frac{R_t}{D_t} + \eta_2 \left(\frac{d_{L_0 - L_N}}{d_{L_t - L_N}} \right) \quad (9)$$

$d_{L_0 - L_N}$ is the distance from the initial point to the terminal location. $d_{L_t - L_N}$ is distance form the current location to the terminal location of the UAV. η_1 and η_2 are weight parameters. They help the agent to capture the effect of data rate and distance. The second part becomes larger when UAV is approaching the terminal location. The objective is to find the optimal policy π^*, among all polices π, that maximizes the sum of the expected rewards, discounted by the factor γ, raised to the power t. The objective function and constrains are defined as:

$$\begin{aligned} (OF): \pi^* &= \underset{\pi}{\operatorname{argmax}} \sum_{t=0}^{T-1} \gamma^t \mathbb{E}_\pi(r_t) \\ C_1: &\quad 0 \le p_{uav} \le p_{max} \\ C_2: &\quad h_{min} \le h_{uav} \le h_{max} \\ C_3: &\quad r_t = 0 \quad if \quad SINR_{uav} \ \& \ SINR_{gue} < Threshold \\ C_4: &\quad r_t = 0 \quad if \quad d_{L_t - L_N} > d_{L_0 - L_N} \end{aligned} \quad (10)$$

γ is in the interval $[0, 1)$. Constraint C_1 requires that the UAV-transmission power p_{uav} be smaller than or equal to p_{max}. Constraint C_2 restricts the UAV altitude between h_{min} and h_{max}, to avoid collision with the ground BS. Constraint C_3 states that the SINR of both ground user (*gue*) and UAV on the shared RB to be greater than or equal to a threshold, otherwise r_t is equal to zero. Constraint C_4 specifies that the distance from the UAV current location L_t to the terminal one L_N is larger than total distance $d_{L_0 - L_N}$, then the reward r_t is zero. It should be noted that at the destination, we have $d_{L_t - L_N} = 0$. Then, the reward is equal to $\eta_1 \cdot \frac{R_t}{D_t}$. In the objective function (*OF*), the BS-agent tries to maximize the cumulative returned reward $\left(\sum_{t=0}^{T-1} r_t \right)$ during the training process, resulting in the minimization of two values (D_t and $d_{L_t - L_N}$). Therefore, it makes a trade-off between minimization of the travelled distance by the UAV and maximization of the users' throughput.

C. DQN Architecture

DQN [16] is a model-free and off-policy method designed for complex environments with many states and actions, consisting of two neural networks, main network and target

network. In this paper, both networks are fully-connected Deep Neural Networks (DNNs) with weight matrices, θ and θ', respectively. The main network acts as a Q-function approximator, mapping a state s_t to all possible actions. Next, the agent employs the ϵ-greedy strategy to select an action a_t. At first, ϵ is set to a value close to one, to conduct exploration of environment. Over time, the ϵ-value is progressively reduced. The agent gradually favors exploitation over exploration. A transition is quadruple from state s_t, action a_t, reward r_t and next state s_{t+1}. All transitions are stored in a buffer called replay memory with capacity $\mathcal{B}$, which is used as a training data set. $\mathcal{B}$ is equal to a set of form $\{(s_1, a_1, r_1, s_2), (s_2, a_2, r_2, s_3), \ldots, (s_t, a_t, r_t, s_{t+1})\}$. When $\mathcal{B}$ contains a number of transitions equal to or greater that the batch size (in this paper, we use 64), then a random mini-batch of transitions is selected from $\mathcal{B}$ to train the main network and update its weight matrix θ. The target network approximates the Q-value for all possible actions of the next state. Its weights are frozen during training. They are updated after M iterations by copying weights of the main network, $\theta' = \theta$. The use of the target network and selecting samples randomly from $\mathcal{B}$ lead to a stable training and reduce the correlation between samples. During training, the goal is to minimize the loss value between outputs of main and target networks using the stochastic gradient descent method, given by:

$$L(\theta) = \frac{1}{P}\sum_{i=1}^{P}(y_i - Q(s_i, a_i; \theta))^2 = \frac{1}{P}\sum_{i=1}^{P}\left[(r_i + \gamma \max_{a'} Q(s_i', a'; \theta')) - Q(s_i, a_i; \theta)\right]^2 \quad (11)$$

where P is the number of samples, $Q(s_i, a_i; \theta)$ is the approximated state-action value of state s_i and action a_i. $y_i = r_i + \gamma \max_{a'} Q(s_i', a'; \theta')$ is the output of the target network. It is the sum of the immediate reward and state-action value for next state s_i'. y_i depends on the type of state s_i'. There are two choices:

$$y_i = \begin{cases} r_i & s_i' \text{ is a terminal state} \\ r_i + \gamma \max_{a'} Q(s_i', a'; \theta') & \text{otherwise} \end{cases}$$

After computing the gradient of the loss function ($\nabla_\theta L(\theta)$), the weights of the main network (θ) are updated using the *Adam* optimizer and learning rate α:

$$\theta_{t+1} = \theta_t - \alpha \nabla_\theta L(\theta_t) \quad (12)$$

This process is repeated until the agent attains a convergence level.

III. Simulation Results

Using simulation with the help of the Pytorch framework, we compare our approach with the OMA, shortest path-planning, and random selection strategies. The single-cell network simulation parameters are listed in Table I. The UAV starts form $L_0 = [220, 200, 80]$ and ends at the destination $L_N = [200, 1100, 120]$. Ground users are assigned a fixed transmission power. Four discrete transmission power values are considered for the UAV. $\mathcal{A}_2$ is equal to $\{p_1, p_2, p_3, p_{max}\}$, which is equal to $\{14, 16, 18, 20\}$, in dBm. The number of available RBs is equal to the number of ground users, i.e., $\mathcal{A}_3$ is equal to $\{RB_1, RB_2\}$. For simplicity, the single-cell cellular-connected UAV network is designed with two ground users. They move randomly with a random speed chosen from interval $[5, 6]$ m/s. The scenario can be generalized to a network with more ground users. With the increase of the number of users in the network, the dimensions of both state space $\mathcal{S}$ and action space $\mathcal{A}$ also increase, requiring a more complex DQN with more layers. Our DQN architecture consists of two 64 neurons hidden layers and one 128 neurons hidden layer with *ReLU* and *Adam* as activation and optimization functions. Fig. 1 plots the loss values for the proposed approach for two threshold values. The loss decreases during training and converges after 17,000 episodes, indicating that the DQN-agent accurately predicts the Q-value of a state and training is stable. Fig. 2 illustrates the received average reward of the NOMA-DQN approach for threshold values of 10 and 15 dB. The increase in reward demonstrates that the agent can successfully learn the dynamics of the environment and design a 3D trajectory for the UAV that maximizes the cumulative reward. However, with a threshold of 15 dB, which induces more inter-user interference (see Fig. 3), the agent initially performs poorly in training. It requires more time to learn the environment and select action pairs, reducing interference. In contrast, the random selection scheme with threshold of 10 dB performs poorly.

Fig. 3 demonstrates the effectiveness of the proposed approach in terms of data rate compared to the orthogonal multiple access (OMA) and shortest path schemes. For the OMA scheme, orthogonal RBs are assigned to users using the OFDM method, reducing the number of sub-action spaces to only $\mathcal{A}_1$. Additionally, the UAV communicates with the ground BS using its maximum transmission power, $p_{\max}$. The proposed framework achieves higher performance in terms of

TABLE I
Simulation Parameters.

Parameters	Value
3D cell size	1500 m × 1500 m × 140 m
Number of ground users	2
Altitude of ground users	1 m
h_{BS}	15 m
Ground users' transmission power	20 dBm
p_{max}, v_{uav}	20 dBm, 20 m/s
η_1, η_2	0.7, 0.3
T	200 s
SINR-thresholds	10, 15 dB
$[h_{min}, h_{max}]$	[80 m, 120 m]
α, ρ_0	3.5, 1.4×10^{-4}
Bandwidth	50 MHz
f_c, N_0	2 GHz, -164 dBm/Hz
Learning rate	1×10^{-5}
Discount factor (γ), Batch size	0.99, 64
Replay memory size ($\mathcal{B}$)	100000
ϵ_{min}, ϵ_{max}, ϵ-decay	0.01, 1, 0.9998

Algorithm 1 NOMA-DQN Learning framework

Input: state space $\mathcal{S}$ and action space $\mathcal{A}$
Initialize replay memory $\mathcal{B}$
Initialize main network Q_θ and target network $Q_{\theta'}$
Initialize target network weights $\theta' \leftarrow \theta$
for $e \leftarrow 1$ to $NumEpisode$ **do**
  Initial state S_0
  $S_0 = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_{uav}, y_{uav}, z_{uav})\}$
  for $t \leftarrow 0$ to $T-1$ **do**
    Generate a random number ($g \in [0, 1]$)
    if $g < \epsilon$ **then**
      $a_t \leftarrow \{a_t^1, a_t^2, a_t^3\}$ is taken randomly
    else
      $a_t = \{a_t^1, a_t^2, a_t^3\} \leftarrow \arg\max(Q(s_t, a_t; \theta))$
    if $d_{L_t - L_N} > d_{L_0 - L_N}$ **then**
      $r_t \leftarrow 0$
    else
      Calculate $SINR_{uav}$ and $SINR_{gue}$
      if $SINR_{uav}$ & $SINR_{gue} \geq Threshold$ **then**
        if $d_{L_t - L_N} \leq d_{L_0 - L_N}$ & $d_{L_t - L_N} \neq 0$ **then**
          $r_t \leftarrow \left[\eta_1 \cdot \frac{R_t}{D_t} + \eta_2 \left(\frac{d_{L_0 - L_N}}{d_{L_t - L_N}}\right)\right]$
        else if $d_{L_t - L_N} = 0$ **then**
          $r_t \leftarrow \left[\eta_1 \cdot \frac{R_t}{D_t}\right]$
      else
        $r_t \leftarrow 0$
    Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{B}$
    if Number of samples in $\mathcal{B} \geq$ Batch size **then**
      Compute loss and update θ
    After M steps, update target network weights: $\theta' \leftarrow \theta$

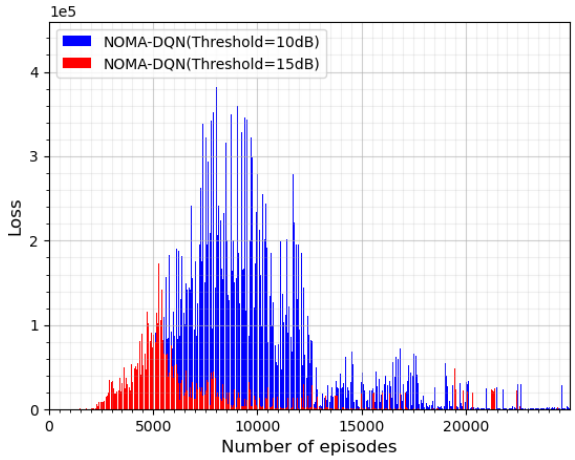

Fig. 1. Loss value vs. number of episodes.

the sum of data rate of aerial and ground users, showcasing the superior spectral efficiency of the NOMA scheme over the OMA method. The shortest path approach is implemented with a threshold of 10 dB. Our approach, also employing a 10 dB threshold, outperforms the shortest path scheme. However, when the threshold is increased to 15 dB, our approach's effectiveness diminishes, resembling to that of the shortest path due to increased inter-user interference. It is noted that the results presented in figures 2 and 3 are smoothed to reduce

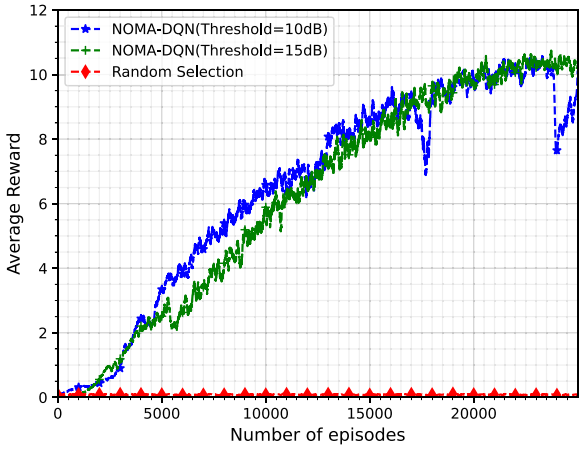

Fig. 2. Average cumulative reward vs. number of episodes.

irregularities in the data, providing a clearer visualization of the learning progress of the agent.

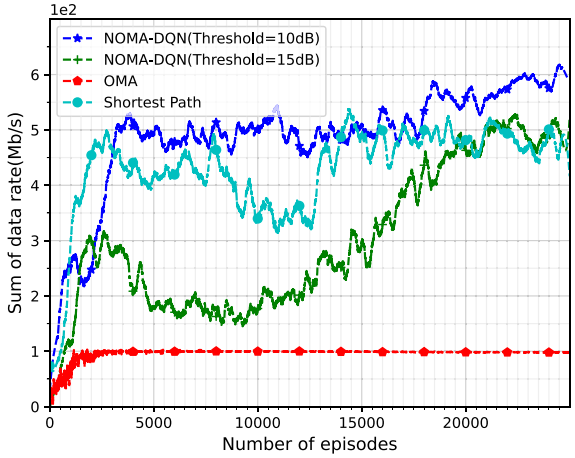

Fig. 3. Sum of data rates of the UAV and ground users vs. number of episodes.

Fig. 4 compares the distance traveled by the UAV from the initial location to the terminal position. The results demonstrate that the proposed optimization function maximizes UAV energy-efficiency by minimizing the number of steps taken by the UAV from its initial location to its target. In essence, the length of the path obtained for NOMA-DQN with a threshold of 15 dB and OMA scheme is equivalent, slightly longer than that of the shortest path scheme, while for NOMA-DQN with a threshold of 10 dB, the UAV travels longer. Conversely, the random selection scheme takes the maximum number of steps without minimizing either the path length or efficiently reaching the UAV destination.

Fig. 5 displays a sample UAV trajectory generated by NOMA-DQN learning for threshold of 15 dB. The efficiency of the path is evaluated using calculating the sum data rate and the UAV traveled distance. At each step, the agent (BS), takes

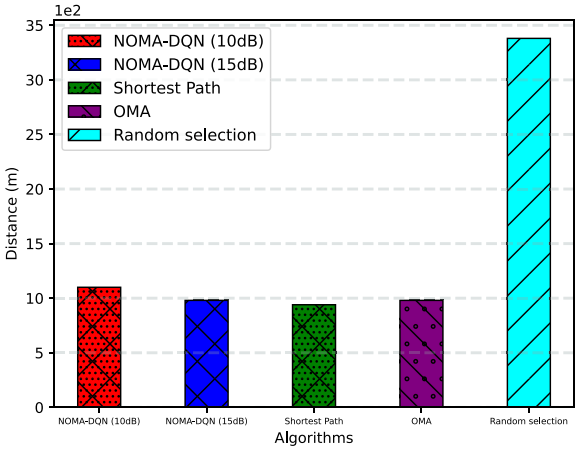

Fig. 4. Distance travelled by the UAV from the initial location to the target.

into account the locations of ground users and UAV. It directs the UAV to a location that minimizes the effect of UAV-uplink interference on its co-channel ground users. It also maximizes the sum data rate of users in the shared RB, until reaching the destination.

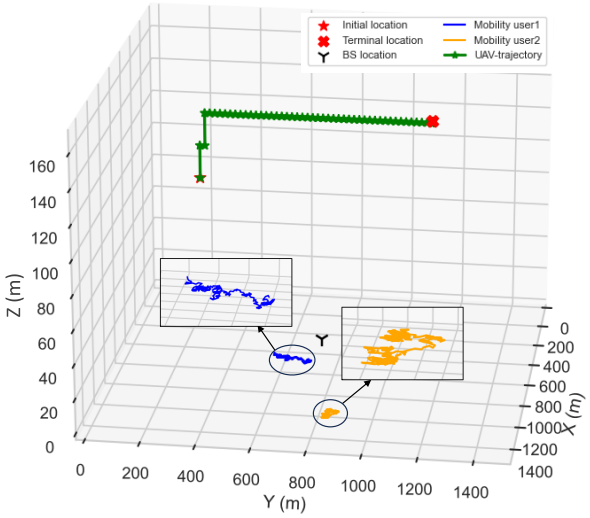

Fig. 5. Mobility of two ground users, and designed 3D UAV trajectory for threshold 15 dB.

IV. Conclusion

We have devised a 3D optimal trajectory for a UAV within a dynamic environment to alleviate the adverse impacts of UAV-uplink interference on ground network users. Simultaneously, our aim was to extend the UAV-battery lifespan by minimizing distance while meeting the minimum data rate requirements for aerial and ground users. Simulation results underscore the superiority of our proposed approach across benchmarks, in terms of total user data rates. Furthermore, our work enhances UAV energy-efficiency by minimizing the number of steps it travels. In the future, we plan to investigate the effectiveness of the proposed approach in real environments, considering existing UAV-energy consumption models.

Acknowledgements — We acknowledge financial support from Ericsson, Mitacs, and NSERC (Natural Sciences and Engineering Research Council) of Canada. The work has also been partially supported by the French National Research Agency under the France 2030 label (NF-HiSec ANR-22-PEFT-0009).

References

[1] Y. Zeng, J. Lyu, and R. Zhang, "Cellular-connected uav: Potential, challenges, and promising technologies," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 120–127, 2018.

[2] D. Mishra and E. Natalizio, "A survey on cellular-connected UAVs: Design challenges, enabling 5G/B5G innovations, and experimental advancements," *Computer Networks*, vol. 182, p. 107451, 2020.

[3] V. Yajnanarayana, Y.-P. E. Wang, S. Gao, S. Muruganathan, and X. L. Ericsson, "Interference mitigation methods for unmanned aerial vehicles served by cellular networks," in *2018 IEEE 5G World Forum (5GWF)*, pp. 118–122, IEEE, 2018.

[4] W. Mei and R. Zhang, "Aerial-ground interference mitigation for cellular-connected UAV," *IEEE Wireless Communications*, vol. 28, no. 1, pp. 167–173, 2021.

[5] X. Pang, G. Gui, N. Zhao, W. Zhang, Y. Chen, Z. Ding, and F. Adachi, "Uplink precoding optimization for NOMA cellular-connected UAV networks," *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 1271–1283, 2020.

[6] W. Mei and R. Zhang, "Uplink cooperative NOMA for cellular-connected UAV," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 644–656, 2019.

[7] A. Zaki-Hindi, R. Amorim, I. Z. Kovács, and J. Wigard, "Uplink coexistence for high throughput UAVs in cellular networks," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pp. 2957–2962, IEEE, 2022.

[8] M. Z. Hassan, G. Kaddoum, and O. Akhrif, "Interference management in cellular-connected internet of drones networks with drone-pairing and uplink rate-splitting multiple access," *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 16060–16079, 2022.

[9] A. Shamsoshoara, F. Lotfi, S. Mousavi, F. Afghah, and I. Guvenc, "Joint path planning and power allocation of a cellular-connected uav using apprenticeship learning via deep inverse reinforcement learning," *arXiv preprint arXiv:2306.10071*, 2023.

[10] P. Li, L. Xie, J. Yao, and J. Xu, "Cellular-connected UAV with adaptive air-to-ground interference cancellation and trajectory optimization," *IEEE Communications Letters*, vol. 26, no. 6, pp. 1368–1372, 2022.

[11] F. Banaeizadeh, M. Barbeau, J. Garcia-Alfaro, V. S. Kothapalli, and E. Kranakis, "Uplink interference management in cellular-connected UAV networks using multi-armed bandit and NOMA," in *2022 IEEE Latin-American Conference on Communications (LATINCOM)*, pp. 1–6, 2022.

[12] A. T. Azar, A. Koubaa, N. Ali Mohamed, H. A. Ibrahim, Z. F. Ibrahim, M. Kazim, A. Ammar, B. Benjdira, A. M. Khamis, I. A. Hameed, *et al.*, "Drone deep reinforcement learning: A review," *Electronics*, vol. 10, no. 9, p. 999, 2021.

[13] F. Bai and A. Helmy, "A survey of mobility models," *Wireless Adhoc Networks. Univ. of Southern California, USA*, vol. 206, p. 147, 2004.

[14] J. B. Andersen, T. S. Rappaport, and S. Yoshida, "Propagation measurements and models for wireless communications channels," *IEEE Communications Magazine*, vol. 33, no. 1, pp. 42–49, 1995.

[15] W. Saad, M. Bennis, M. Mozaffari, and X. Lin, *Wireless Communications and Networking for Unmanned Aerial Vehicles*. Cambridge University Press, 2020.

[16] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, 2020.

[17] W. K. New, C. Y. Leow, K. Navaie, Y. Sun, and Z. Ding, "Application of NOMA for cellular-connected UAVs: Opportunities and challenges," *Science China Information Sciences*, vol. 64, no. 4, pp. 1–14, 2021.