IRS Empowered UAV Wireless Communication with Resource Allocation, Reflecting Design and Trajectory Optimization

Xiaoqi Zhang, Haijun Zhang, *Senior Member, IEEE*, Wenbo Du, *Member, IEEE*, Keping Long, *Senior Member, IEEE*, Arumugam Nallanathan, *Fellow, IEEE*

Abstract

As revolutionary technologies that can actively change the communication link signal, intelligent reflecting surface (IRS) and unmanned aerial vehicle (UAV) have emerged as reliable, economical and convenient wireless communication solutions for a variety of practical scenarios. Therefore, this paper focuses on an IRS empowered UAV downlink communication network, where the dynamic UAV establishes a cascade link via IRS to provide signal enhancement services for multiple users. Considering constraints of transmit power, flight speed and area at the UAV and the reflecting constraints at the IRS, the block coordinate descent (BCD) method based on resource allocation, reflecting design and trajectory optimization is adopted to maximize the sum-rate of all users. The proposed problem is converted by using quadratic transformation and Lagrangian dual transformation. Then applying for the approximate linear method and Iterative Rank Minimization (IRM) to optimize the transmit power of UAV and phase shift of IRS respectively. Since additional reflection propagation paths by IRS, the complexity of the channel model makes the trajectory design difficult. To tackle this problem, this paper proposes a UAV trajectory optimization method based on enhanced reinforcement learning with the fixed initial location and destination. In the end, the convergence of the proposed scheme is effectively verified

X. Zhang, H. Zhang, and K. Long are with the Institute of Artificial Intelligence, Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, University of Science and Technology Beijing, Beijing 100083, China (e-mail: zhangxiaoqi@xs.ustb.edu.cn; haijunzhang@ieee.org; longkeping@ustb.edu.cn).

W. Du is with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: wenbodu@buaa.edu.cn).

A. Nallanathan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: a.nallanathan@qmul.ac.uk).

by simulations. Moreover, abundant simulation comparisons between the proposed scheme and other benchmark schemes demonstrate the validity and high performance gains of the proposed algorithm.

Index Terms

IRS, UAV, resource allocation, reflecting design, trajectory optimization, BCD, IRM, reinforcement learning.

I. Introduction

In recent years, UAV and IRS have become two promising technologies to facilitate the development of wireless communication networks by actively changing communication link signals through maneuver control and intelligent signal reconstruction respectively [1] [2]. Nevertheless, due to their limitations and great challenges in practice, their future applications have been seriously hindered [3]. To meet the strict service requirements of future networks, it is significant to make full use of the complementary advantages of IRS and UAV [4] [5].

A. Related Works and Motivation

With the gradual maturity of UAV technology and the decreasing cost, UAV communication is widely employed in environmental monitoring, agricultural production, news reporting, film shooting, electrical inspection and other fields. On account of the agility and maneuverability of UAVs, UAVs can be quickly deployed in target areas to establish reliable communication links [6]. Moreover, the high mobility, flexible deployment and agility of UAVs enable it to be deployed quickly and effectively establish on-demand communications in emergency situations [7]. Specifically, as a complement to the existing wireless network, UAVs can provide additional capabilities for hotspots and provide coverage of remote areas in poor conditions [8]. In addition, when an emergency occurs, the UAV base station is not limited by the basic communication facilities, and can quickly provide a large range of reliable communications for the disaster area [9]. Compared with traditional communication infrastructure, the deployment of UAVs is more affordable. Utilizing the UAV communications to improve wireless network coverage is a cost-effective choice [10].

At present, the trajectory optimization and placement design in the UAV communication network have received extensive attention [11]. For example, in [12], the authors established

a system network of UAV wireless power transmission, where the asymptotic optimal solution of UAV trajectory design is derived. In [13], an alternate optimization algorithm for maximizing energy efficiency based on resource allocation and trajectory optimization is proposed. Furthermore, the authors proposed a dual UAV system consisting of a communication UAV and a jamming UAV, where achieved the maximization of the secrecy energy efficiency based on resources, trajectory and artificial noise through successive convex approximations [14]. In [15], the authors focused on a UAV-assisted wireless power transfer network while considering the nonlinear energy harvesting process. In the case of the maximum speed limit of the UAV, the trajectory design aiming at maximizing the minimum capture energy between ground equipment was studied. In [16], the authors concentrated on the multi-UAV Internet of Things (IoT) system under the uplink NOMA communication. Through reasonable design of UAV flight height, the sub-channel allocation, uplink transmit power and device node are jointly designed to maximize the system capacity. To minimize the energy consumption, authors of [17] proposed to design the UAV placement while meeting the system throughput requirements of each user. The authors in [18] optimized the 3D trajectory of the UAV by leveraging the proximal difference-of-convex algorithm with extrapolation method, and extended to an online optimization. In [19], Capitalizing on the successive convex approximation method, the author proposed an effective iterative method to find feasible solutions that satisfy the Karush-Kuhn-Tucker condition, and achieved the goal of optimizing the trajectory of the UAV to minimize the energy consumption.

However, while the development of UAVs' communication is promising, they have strict size, energy constraints and load limits which make their flight time or endurance difficult, as well as their communication performance [3]. In addition, for densely built-up areas, it is usually necessary to rasie the altitude of the UAV and the transmit power of the base station to construct a line-of-sight (LOS) connection with users. However, this operation usually leads to greater path loss and energy loss, and physical signal blocking and interference of tall buildings and other obstacles will lead to frequent occurrence of high packet loss rate during transmission.

As an emerging wireless transmission technology of 6G, IRS can be used to complement and enhance the quality of signal transmission in wireless communication networks [20]. Specially, IRS has excellent features of portable and low-cost. By constructing an intelligent and controllable wireless environment, IRS will bring a new communication network paradigm to 6G and meet future wireless communication needs. The simplified version of IRS will have

the opportunity for initially commercial deployment and standardization in the 5G-Advanced stage, especially to improve the 5G millimeter wave coverage problem [21], [22]. The authors of [23] studied a communication system based on IRS-assisted SWIPT under QoS constraints by jointly designing active and passive beamforming. In [24], the author jointly optimized the minimum total transmit power of the system by designing active and passive beamforming, significantly improving network energy consumption and increasing the achievable rate. The paper [25] developed an energy-saving design through IRS phase design and transmit power optimization. Aiming at imperfect channel information, a novel algorithm based on penalty binary decomposition was proposed to achieve the amplitude control of the IRS in wireless communication [26]. Consider perfect setting of channel state information (CSI) as well as imperfect CSI, the joint design of the AP beamforming and the reflecting phase shift of IRS is separately studied to make the weighted sum-rate of multiusers maximize [27]. To maximize the security rate, considering various QoS requirements, the problem of joint beamforming and reflection beamforming and numerical analysis was carried out with deep reinforcement learning by the authors of [28]. In [29], a machine learning method with implicit channel estimation was proposed, which can directly optimize the reflection coefficient of the IRS.

Although the combination of IRS and UAV is in its infancy, some of their researches have attracted great attentions. IRS empowered UAV communication network can not only enhance the reliability of the communication link, but also improve the system performance gain. When the user or the IoT device is in a blind spot, IRS establishes line-of-sight links through intelligent reflection to bypass obstacles and solve the problem of signal coverage dead zone. In view of the above advantages, through joint optimization phase shift and UAV scheduling and trajectory, the weighted bit error rate minimization was explored in [4]. The authors of [30] investigated the joint design of resource allocation, beamforming and UAV placement of IoT devices under the constraint of finite block length. By means of jointly designing the speed, trajectory of each UAV and phase shift, the goal of minimizing the total transmit power was studied in [31]. In [32], the authors considered the worst-case secrecy rate minimum through a reliable joint design of beamforming, UAV's transmit power as well as trajectory. In [33], the authors studied the joint optimization of UAV trajectory, transmit beamforming and IRS passive beamforming to maximize the average achievable rate of the relay network by equipping the IRS on a UAV. By optimizing the UAV movement and IRS design, the energy consumption minimization problem

was proposed in [34].

According to the above investigations, although researchers have done a handful of researches on the IRS-aided UAV communication system, this research is still in a fledging period. As far as we know, the optimization problem of IRS empowered UAV wireless network under the maximum sum-rate requirements has not been solved yet, and it is still an extremely attractive research. As the IRS-assisted UAV communication network has been proven by several researchers to bring excellent performance improvements, actively exploring new optimization methods to carry out this work is our motivation. Different from the above research, the reflecting design algorithm of the IRS employed in this paper is first used in the related research of UAV wireless networks. In addition, the trajectory optimization algorithm proposed in this paper is also a groundbreaking work in the scenario involving IRS.

B. Contributions

This paper investigates an IRS empowered UAV wireless communication network considering the optimization problem of joint resource allocation, reflecting design and trajectory optimization to maximum sum-rate of multiusers. The main contributions are listed below.

- First of all, this paper proposes an IRS empowered UAV system communication model, in which the UAV performs flight tasks at specified initial location and destination within a certain area and the IRS is fixed to a tall building. The system model can not only takes full advantage of the high mobility of the UAV, but also provides more reliable, individualized communication services for multiusers by virtue of IRS enhanced links. Based on the transmit power limits and the reflecting constraints of the IRS, a joint optimization problem of resource allocation, IRS reflecting design and UAV trajectory optimization to maximize the sum-rate of multiusers is proposed.
- Secondly, this paper exploits an iterative method based on BCD, which decomposes the complexly coupled joint optimization problem into three subproblems. For the first subproblem, the problem of power control is transformed through quadratic transformation and Lagrangian dual transformation, and approximate linear algorithm is employed to solve the proposed subproblem. While an IRM method is effectively applied to solve the second subproblem for the reflecting design of IRS. Due to the deployment of IRS introduces

additional reflection propagation paths, to tackle this problem, an enhanced reinforcement learning method is employed to carry out the third subproblem skillfully.

- Finally, abundant simulations unveil the remarkable performance of the proposed joint optimization scheme, and the proposed scheme converges with fewer iterations. Compared with several benchmark schemes, the scheme can markedly promote the sum-rate of multiusers and prove the superiority of the scheme. Moreover, the deployment of IRS has a significant effect on improving the communication quality. The reflecting numbers and phase shift design of IRS can also greatly promote the sum-rate. What is more, the reinforcement learning algorithm for trajectory optimization in the joint problem plays a notable role.

C. Organization and Notations

The remainder of this paper is organized as follows. Section II first introduces the system model of IRS empowered UAV downlink communications network. Based on the system model and channel model, the problem of maximizing the cumulative sum-rate of all the users is formulated. In Section III, a BCD algorithm is proposed to solve the joint design problem. Section IV provides plentiful simulation analysis, which are compared with other benchmark schemes to demonstrate the validity of the scheme. In the end, Section V makes a conclusion for this paper.

Notation: $\mathbf{b}^*$ and $\mathbf{b}^H$ mean the transpose and conjugate transpose of vector $\mathbf{b}$ respectively. $\|\mathbf{b}\|$ represents the Euclidean norm. $trace(\mathbf{B})$ and $rank(\mathbf{B})$ represent the trace and rank of matrix $\mathbf{B}$, respectively. $\mathbf{B} \succeq 0$ denotes $\mathbf{B}$ as a positive semi-definite matrix. $|x|$ is the absolute value of the complex number x, while $\mathrm{Re}\{x\}$ is the real part. $\langle \cdot \rangle$ represents the inner product of two matrices.

II. System Model and Problem Formulation

In this section, the system model and channel model of the IRS empowered UAV downlink communications network are presented, and then the joint optimization problem is formulated under multiple constraints.

A. System Model

Fig. 1 exhibits a dynamic UAV which assumes the role of a base station in the air, and it will provide downlink communication services for K single-antenna user equipment (UE) in a

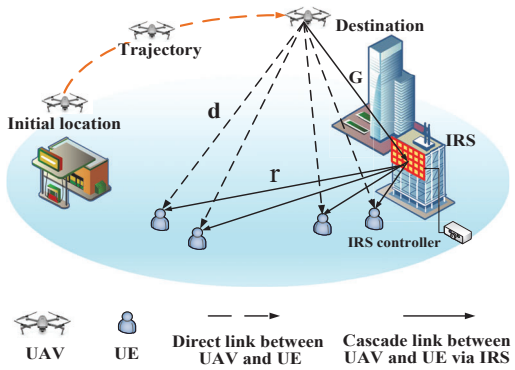


Fig. 1. IRS empowered UAV downlink communications network.

certain area. It is assumed that the movement of UE is static or low mobility, and the horizontal location of UE i is expressed as $\mathbf{u}_i = [x_i^{UE}, y_i^{UE}]^T$, $i \in K$. In addition to the direct link between UAV and UE, the system establishes a cascade link to enhance the UE's received signal with the aid of M reflecting elements in IRS. In general, the IRS are deployed on the tall buildings with good sight conditions. Therefore, the IRS is fixed in a specific location and represented as $\mathbf{I} = [x^{IRS}, y^{IRS}, h^{IRS}]^T$.

To maintain the consistency of UAV flight time, the total flight time T is divided into N equal time intervals. The 3D location of the UAV in time slot n can be expressed as $\mathbf{q}[n] = [x^{UAV}[n], y^{UAV}[n], h^{UAV}]^T$, $n \in N$, and h^{UAV} denote the flying altitude of the UAV. For simplicity, the height is assumed to be constant in this paper. The task of the UAV is to fly between a predetermined set of initial location and destination, which are denoted by $\mathbf{q}_I$ and $\mathbf{q}_F$ respectively.

$$\mathbf{q}[0] = \mathbf{q}_I, \mathbf{q}[N] = \mathbf{q}_F. \tag{1}$$

In addition to flying in accordance with the initial location and destination, it is assumed that there is a no-fly zone restriction, which is expressed as (2). It is worth mentioning that the no-fly zone is interpreted as an area where obstacles exist at the height of the UAV (such as a tall building), or an area listed as a no-fly zone by regulatory affairs.

$$\mathbf{q}[n] \in C \backslash C^{no-fly}, \forall n \in N. \tag{2}$$

In each time slot, even if the UAV reaches the maximum speed $V_{\max}$, as long as the time slots $\delta_n = \frac{T}{N}$ is divided into sufficiently small, the location of the UAV can be regarded as

approximately constant. Taking into account the maximum speed $V_{\max}$ which the UAV can achieve, the UAV mobility constraints in a time slot is as follows

$$\|\mathbf{q}[n] - \mathbf{q}[n-1]\| \leq \frac{V_{\max} T}{N}, \forall n \in N. \tag{3}$$

In actual application scenarios, the IRS is usually equipped with an intelligent controller to configure the reflection coefficient and exchange information between the IRS and the UAV. Due to the large channel fading of the UAV-IRS-UE link, the reflective link requires numerous reflective components to make up path loss. The reflection coefficient of the IRS is regarded as approximately unchanged over the whole signal bandwidth. Then the IRS reflecting matrix is expressed as $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_m, \cdots, \theta_M]^H$, where $\theta_m = e^{j\varphi_m}$, φ_m is the reflecting phase shift of the m-th unit corresponding to IRS. The distance of UAV-UE(i) and UAV-IRS in time slot $n$is respectively expressed as

$$D_i^{UE}[n] = \|\mathbf{q}[n] - \mathbf{u}_i\|, \tag{4}$$

$$D^{UR}[n] = \|\mathbf{q}[n] - \mathbf{I}\|. \tag{5}$$

Since the movement of the UAV during δ is far less than $D^{UR}[n]$ and $D_i^{UE}[n]$, $D^{UR}[n]$ and $D_i^{UE}[n]$ are constant in each time slot n. Moreover, the distance of IRS-UE(i) is expressed as (6), which can be treated as fixed in the model.

$$D_i^{RE} = \|\mathbf{I} - \mathbf{u}_i\|. \tag{6}$$

In virtue of the substantial path fading and reflection loss, signals that are reflected twice or more by IRS can be ignored [2].

B. Channel Model

In this paper, $\mathbf{d}_i[n]$ represents the equivalent baseband channel vector of UAV-UE link in the time slot n of the IRS empowered UAV downlink communication network. In time slot n, the equivalent baseband channel vector of IRS-UE link is denoted by $\mathbf{r}_i[n]$. $\mathbf{G}[n]$ represents the channel gain vector of the UAV-IRS link. Since UAVs fly in the air with a certain height, IRS is usually deployed at tall buildings to avoid signal blocking, as shown in Fig. 1. It is assumed that the channels $\mathbf{d}_i[n]$ and $\mathbf{r}_i[n]$ follows the Rician channel model. Considering large-scale and

small-scale fading, the path loss is composed of LOS and non-line-of-sight (NLOS) link, which can be respectively represented as

$$\mathbf{d}_i[n]=\sqrt{\frac{\rho_0}{(D_i^{UE}[n])^{\alpha^d}}}\left(\sqrt{\frac{\kappa^d}{\kappa^d+1}}\overline{\mathbf{d}}_i[n]+\sqrt{\frac{1}{\kappa^d+1}}\overline{\overline{\mathbf{d}}}_i[n]\right), \tag{7}$$

$$\mathbf{r}_i[n]=\sqrt{\frac{\rho_0}{(D_i^{RE})^{\alpha^r}}}\left(\sqrt{\frac{\kappa^r}{\kappa^r+1}}\overline{\mathbf{r}}_i+\sqrt{\frac{1}{\kappa^r+1}}\overline{\overline{\mathbf{r}}}_i[n]\right), \tag{8}$$

where ρ_0 is the path loss when the reference distance is 1 meter. κ^d and κ^r are Rician factors. α^d and α^r are the path loss exponents of UAV-UE and IRS-UE links. $\overline{\mathbf{d}}_i[n]=1$ and $\overline{\mathbf{r}}_i$ denote deterministic LOS components, $\overline{\overline{\mathbf{d}}}_i[n]$ and $\overline{\overline{\mathbf{r}}}_i[n]$ are random Rayleigh distribution NLOS components, which conform to cyclic symmetric complex Gaussian (CSCG) distribution with zero mean and unit variance.

It is assumed that the IRS consists of uniform linear array (ULA) reflection elements, which $\overline{\mathbf{r}}_i$ can be represented as

$$\overline{\mathbf{r}}_i=\left[1,e^{-j\frac{2\pi d}{\lambda}\cos\phi_i},...,e^{-j\frac{2\pi(M-1)d}{\lambda}\cos\phi_i}\right]^T, \tag{9}$$

where d is the IRS unit spacing, λ represents the carrier wavelength, $\cos\phi_i=\frac{x_i^{UE}-x^{IRS}}{D_i^{RE}}$ denotes the cosine of the angle of departure (AoD) from the IRS to UE i.

Moreover, the equivalent baseband channel vector of UAV-IRS link can be denoted as

$$\mathbf{G}[n]=\sqrt{\frac{\rho_0}{(D^{UR}[n])^2}}\overline{\mathbf{G}}_i[n], \tag{10}$$

$$\overline{\mathbf{G}}_i[n]=\left[1,e^{-j\frac{2\pi d}{\lambda}\cos\varphi[n]},...,e^{-j\frac{2\pi(M-1)d}{\lambda}\cos\varphi[n]}\right]^T, \tag{11}$$

where $\cos\varphi[n]=\frac{x^{IRS}-x^{UAV}[n]}{D^{UR}[n]}$ denotes the cosine of the angle of arrival (AoA) from the UAV to the IRS.

Let $s_i[n]$ represents the data symbol sent to UE i in time slot n, and $s_i[n]$ is an independent random variable with zero mean and unit variance. Furthermore, the transmit power of UAV satisfies $\sum_{i=1}^{K}\sum_{n=1}^{N}\mathbf{p}_i[n]\leq P_{UAV}$, where $\mathbf{p}_i[n]$ denotes the transmit power vector of the UAV at UE i in time slot n, P_{UAV} represents the maximum transmit power of the UAV. Thus the transmitted signal at the UAV can be represented as

$$\mathbf{x}[n]=\sum_{i=1}^{K}\sqrt{\mathbf{p}_i[n]}s_i[n]. \tag{12}$$

Compared with the traditional communication network, the UE received signal in this paper is composed of two parts, namely signals of direct link (UAV-UE link) and reflecting link (UAV-IRS-UE link). Therefore, the ith UE receives the signal in time slot n as

$$y_i(n) = \underbrace{\mathbf{d}_i^H[n]\sqrt{\mathbf{p}_i[n]}s_i[n]}_{\mathrm{UAV-UElink}} + \underbrace{\boldsymbol{\theta}^H[n]\mathbf{R}_i[n]\sqrt{\mathbf{p}_i[n]}s_i[n]}_{\mathrm{UAV-IRS-UElink}} + \tau_i, \tag{13}$$

where $\mathbf{R}_i[n]{=}diag(\mathbf{r}_i^H[n])\mathbf{G}[n]$. τ_i represents the additive white Gaussian noise (AWGN) with zero mean and variance σ_0^2. Then the corresponding received signal-to-noise ratio (SINR) of UE i in time slot n is represented as

$$SINR_i\left[n\right] = \frac{\left|\left(\mathbf{d}_i^H\left[n\right] + \boldsymbol{\theta}^H\left[n\right]\mathbf{R}_i\left[n\right]\right)\right|^2\mathbf{p}_i\left[n\right]}{\sum\limits_{j=1,j\neq i}^{K}\left|\left(\mathbf{d}_i^H\left[n\right] + \boldsymbol{\theta}^H\left[n\right]\mathbf{R}_i\left[n\right]\right)\right|^2\mathbf{p}_j\left[n\right] + \sigma_0^2}. \tag{14}$$

The data rate of UE i in time slot n is represented as

$$R_i[n] = \log(1 + SINR_i[n]). \tag{15}$$

Since the optimal solution has nothing to do with the base of the logarithmic function, the natural logarithm is used.

C. Problem Formulation

In this subsection, the goal of maximizing the sum-rate of total UE throughout the flight of the UAV is proposed. Then, convert the multiple variable optimization problem considered in this paper to the following formula

$$\begin{aligned} P:\ &\max_{\mathbf{P},\boldsymbol{\theta},\mathbf{Q}} f(\mathbf{P},\boldsymbol{\theta},\mathbf{Q}) = \sum_{i=1}^{K}\sum_{n=1}^{N} R_i[n] \\ s.t\quad &C1: \sum_{i=1}^{K}\sum_{n=1}^{N}\mathbf{p}_i[n] \le P_{UAV} \\ &C2: \left|\theta_m[n]\right| = 1, \forall m = 1,...,M \\ &C3: \left\|\mathbf{q}[n] - \mathbf{q}[n-1]\right\| \le \frac{V_{\max}T}{N}, \forall n \in N \\ &C4: \mathbf{q}[0] = \mathbf{q}_I, \mathbf{q}[N] = \mathbf{q}_F \\ &C5: \mathbf{q}[n] \in C\backslash C^{no-fly}, \forall n \in N \end{aligned}, \tag{16}$$

where $\mathbf{P}{=}\{\mathbf{p}_i, \forall i \in K\}$ denotes the UAV transmit power, $\mathbf{Q}{=}\{\mathbf{q}[n], \forall n \in N\}$ is the trajectory coordinate of the UAV, $\boldsymbol{\theta}{=}\{\theta_m, \forall m \in M\}$ represents the IRS phase shift matrix. $C1$ represents the

Algorithm 1 BCD-based for solving problem P

1: Initialize $\{\mathbf{P}^0, \boldsymbol{\theta}^0, \mathbf{Q}^0\}$, and set $t = 0$.
2: **repeat**
3: Solve PA for given $\{\boldsymbol{\theta}^t, \mathbf{Q}^t\}$ by employing Algorithm 2, and update $\mathbf{P}^{t+1}$.
4: Solve PB for given $\{\mathbf{P}^{t+1}, \mathbf{Q}^t\}$ by employing Algorithm 3, and update $\boldsymbol{\theta}^{t+1}$.
5: Solve PC for given $\{\mathbf{P}^{t+1}, \boldsymbol{\theta}^{t+1}\}$ by employing Algorithm 4, and update $\mathbf{Q}^{t+1}$.
6: Update $t = t + 1$.
7: **until** $t = t_{\max}$

UAV transmit power constraint, $C2$ represents the IRS phase shift constraint, and $C3$ guarantees that the UAV is equivalent to the approximately constant location in a time slot, $C4$ specifies the initial location and destination of the UAV. $C5$ denotes that the UAV is not allowed to fly over the no-fly zone.

There are three unknown variables in the system model. Obviously, the objective function and the multiple constraints are non-convex problems. Therefore, the above problems are NP-hard and hard to carry out. To address this problem, an iterative algorithm is adopted to find a sub-optimal solution by invoking the BCD method in the next section .

III. Proposed Scheme

In this section, the BCD method [35] is used to alternately optimize the three subproblems. Specifically, the variables that need to be optimized are decomposed into several blocks, other block parameters are fixed, and each block is updated according to specific rules. The detailed process is as follows.

Firstly, for given IRS phase shift $\boldsymbol{\theta}$ and UAV trajectory $\mathbf{Q}$, UAV transmit power $\mathbf{P}$ allocation is optimized based on an approximate linear algorithm by solving a linear programming problem. Secondly, fix $\mathbf{P}$ and $\mathbf{Q}$, $\boldsymbol{\theta}$ is optimized based on IRM algorithm. Finally, an enhanced reinforcement learning method is adopted to optimize the UAV trajectory. The joint optimization process is as algorithm 1.

A. Resource Allocation

As for the solution of the resource allocation in IRS aided UAV network, the problem is transformed through the Lagrange dual transformation and quadratic transformation [36], and an approximate linear method is applied to solve the transmit power control of UAV. Therefore,

fixing the phase shift of the IRS and the trajectory of the UAV, the problem of power control can be transformed into P(A) as follows

$$\begin{aligned} P(A): \max_{\mathbf{P}} & f_A(\mathbf{P}) = \sum_{i=1}^{K}\sum_{n=1}^{N} R_i[n] \\ s.t \quad & \sum_{i=1}^{K}\sum_{n=1}^{N} \mathbf{p}_i[n] \le P_{UAV} \end{aligned} . \tag{17}$$

It can be seen that the above formula is a multidimensional and complex fractional programming (FP) problem [36]. To trakle the problem, the closed-form FP method is applied to equivalently convert the logarithm of the ratio problem into a more manageable form. According to the Lagrangian dual transformation proposed in [37], the objective function f_A is as follows

$$f_{A1}(\mathbf{P}, \boldsymbol{\mu}) = \sum_{i=1}^{K}\sum_{n=1}^{N} \log(1+\mu_i[n]) - \sum_{i=1}^{K}\sum_{n=1}^{N} \mu_i[n] + \sum_{i=1}^{K}\sum_{n=1}^{N} \frac{(1+\mu_i[n])SINR_i[n]}{1+SINR_i[n]} , \tag{18}$$

where the auxiliary variable $\boldsymbol{\mu}$ refers to $[\mu_1, ..., \mu_i ..., \mu_K]^T$, $\mu_i[n] \ge 0, \forall i = 1, ..., K$.

Aiming at the multiple-ratio FP problem in formula (18), the quadratic transformation proposed in [37] is employed to transform the problem. Formula (18) can be rewritten as a biconvex optimization problem in (19), where the auxiliary variables $\boldsymbol{\omega}$ refers to $[\omega_1, ..., \omega_i ..., \omega_K]^T$.

$$\begin{aligned} f_{A2}(\mathbf{P}, \boldsymbol{\mu}, \boldsymbol{\omega}) = & \sum_{i=1}^{K}\sum_{n=1}^{N} \left(\log(1+\mu_i[n]) - \mu_i[n]\right) \\ & + \sum_{i=1}^{K}\sum_{n=1}^{N} 2\sqrt{(1+\mu_i[n])\mathbf{p}_i[n]}\mathrm{Re}\{\omega_i^*[n](\mathbf{d}_i^H[n] + \boldsymbol{\theta}^H[n]\mathbf{R}_i[n])\} \\ & - \sum_{i=1}^{K}\sum_{n=1}^{N} |\omega_i[n]|^2 (\sum_{j=1}^{K}\sum_{n=1}^{N} \left|(\mathbf{d}_i^H[n] + \boldsymbol{\theta}^H[n]\mathbf{R}_i[n])\right|^2 \mathbf{p}_j[n] + \sigma_0^2) \end{aligned} . \tag{19}$$

Then problem P(A) can be rewritten as

$$\begin{aligned} \widehat{P(A)}: \max_{\mathbf{P}, \boldsymbol{\mu}, \boldsymbol{\omega}} & f_{A2}(\mathbf{P}, \boldsymbol{\mu}, \boldsymbol{\omega}) \\ s.t \quad & \sum_{i=1}^{K}\sum_{n=1}^{N} \mathbf{p}_i[n] \le P_{UAV} \\ & \mu_i[n] \ge 0, \forall i = 1, ..., K \end{aligned} . \tag{20}$$

For the problem (20), the variables $\mathbf{p}$, $\boldsymbol{\mu}$, $\boldsymbol{\omega}$ are updated circularly through iteration and alternation. The specific update process by the approximate linear method is as algorithm 2. Considering the parameters $\mathbf{p}$ and $\boldsymbol{\omega}$ as constants, solve the formula $\partial f_{A2}/\partial \mu_i[n]=0$ as follows

$$\frac{\partial f_{A2}}{\partial \mu_i[n]} = \sum_{i=1}^{K}\sum_{n=1}^{N} \left(\frac{-\mu_i[n]}{1+\mu_i[n]} + \frac{\sqrt{\mathbf{p}_i[n]}\mathrm{Re}\{\omega_i^*[n](\mathbf{d}_i^H[n] + \theta^H[n]\mathbf{R}_i[n])\}}{\sqrt{(1+\mu_i[n])}} \right) ,$$

let $\partial f_{A2}/\partial \mu_i[n]=0$, the following equation can be obtained

$$\frac{\mu_i[n]}{1+\mu_i[n]}=\frac{\sqrt{\mathbf{p}_i[n]}\mathrm{Re}\{\omega_i^*[n](\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])\}}{\sqrt{(1+\mu_i[n])}}$$

let $\chi_i[n]=\sqrt{\mathbf{p}_i[n]}\mathrm{Re}\{\omega_i^*[n](\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])\}$. Then solve equation $\mu_i^2[n]-\chi_i^2[n]\mu_i[n]-\chi_i^2[n]=0$ to get $\mu_i[n]$ as follow

$$\mu_i[n]=\frac{\ddot{\chi}_i^2[n]+\ddot{\chi}_i[n]\sqrt{\ddot{\chi}_i^2[n]+4}}{2}\ , \tag{21}$$

where $\ddot{\chi}$ represents the updated value of the parameters χ.

Similarly, solve $\partial f_{A2}/\partial \omega_i[n]=0$ in the following

$$\begin{aligned}\frac{\partial f_{A2}}{\partial \omega_i[n]}&=\sum_{i=1}^{K}\sum_{n=1}^{N}\sqrt{(1+\mu_i[n])\mathbf{p}_i[n]}\frac{\partial(\omega_i^H[n](\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])+(\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])^H\omega_i[n])}{\partial\omega_i[n]}\\&-\sum_{i=1}^{K}\sum_{n=1}^{N}\frac{\partial(\omega_i[n]\omega_i^H[n])}{\partial\omega_i[n]}(\sum_{j=1}^{K}\sum_{n=1}^{N}\left|(\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])\right|^2\mathbf{p}_j[n]+\sigma_0^2)\\&=\sum_{i=1}^{K}\sum_{n=1}^{N}\sqrt{(1+\mu_i[n])\mathbf{p}_i[n]}(\mathbf{0}+(\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])^H)\\&-\sum_{i=1}^{K}\sum_{n=1}^{N}\omega_i^H[n](\sum_{j=1}^{K}\sum_{n=1}^{N}\left|(\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])\right|^2\mathbf{p}_j[n]+\sigma_0^2)\end{aligned},$$

let $\partial f_{A2}/\partial \omega_i[n]=0$, the following equation can be obtained

$$\begin{aligned}&\sum_{i=1}^{K}\sum_{n=1}^{N}\omega_i^H[n](\sum_{j=1}^{K}\sum_{n=1}^{N}\left|(\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])\right|^2\mathbf{p}_j[n]+\sigma_0^2)\\&=\sum_{i=1}^{K}\sum_{n=1}^{N}\sqrt{(1+\mu_i[n])\mathbf{p}_i[n]}(\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])^H\end{aligned},$$

then update $\omega_i[n]$ as

$$\omega_i[n]=\frac{\sqrt{\ddot{\mathbf{p}}_i\left[n\right]\left(1+\ddot{\mu}_i[n]\right)}(\mathbf{d}_i^H\left[n\right]+\ddot{\boldsymbol{\theta}}^H\left[n\right]\mathbf{R}_i\left[n\right])}{\sum_{j=1}^{K}\left|(\mathbf{d}_i^H\left[n\right]+\ddot{\boldsymbol{\theta}}^H\left[n\right]\mathbf{R}_i\left[n\right])\right|^2\ddot{\mathbf{p}}_j\left[n\right]+\sigma_0^2}. \tag{22}$$

Introducing the dual vector β constrained by the UAV transmit power, then the dual function of f_{A2} is

$$\begin{aligned}f_{A2}^D(\mathbf{P},\mu,\omega)&=\sum_{i=1}^{K}\sum_{n=1}^{N}(\log(1+\mu_i[n])-\mu_i[n])\\&+\sum_{i=1}^{K}\sum_{n=1}^{N}2\sqrt{(1+\mu_i[n])\mathbf{p}_i[n]}\mathrm{Re}\{\omega_i^*[n](\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])\}\\&-\sum_{i=1}^{K}\sum_{n=1}^{N}|\omega_i[n]|^2(\sum_{j=1}^{K}\sum_{n=1}^{N}\left|(\mathbf{d}_i^H[n]+\theta^H[n]\mathbf{R}_i[n])\right|^2\mathbf{p}_j[n]+\sigma_0^2)\\&-\beta\left(\sum_{i=1}^{K}\sum_{n=1}^{N}\mathbf{p}_j[n]-P_{UAV}\right)\end{aligned}.$$

Decompose f_{A2}^{D} as follows

$$f_{A2}^{D} = const(\mu_i[n]) + f_1 + f_2 + f_3,$$

$$f_1 = \sum_{i=1}^{K}\sum_{n=1}^{N} 2\sqrt{(1+\mu_i[n])\mathbf{p}_i[n]}\mathrm{Re}\{\omega_i^*[n](\mathbf{d}_i^H[n] + \theta^H[n]\mathbf{R}_i[n])\},$$

$$f_2 = -\sum_{i=1}^{K}\sum_{n=1}^{N} |\omega_i[n]|^2 (\sum_{j=1}^{K}\sum_{n=1}^{N} \left|(\mathbf{d}_i^H[n] + \theta^H[n]\mathbf{R}_i[n])\right|^2 \mathbf{p}_j[n] + \sigma_0^2),$$

$$f_3 = -\beta\left(\sum_{i=1}^{K}\sum_{n=1}^{N}\mathbf{p}_j[n] - P_{UAV}\right),$$

the partial derivative can be obtained as follows

$$\frac{\partial f1}{\partial\sqrt{\mathbf{p}_i[n]}} = \sum_{i=1}^{K}\sum_{n=1}^{N}\sqrt{(1+\mu_i[n])}\left[\left(\mathbf{d}_i[n] + \theta[n]\mathbf{R}_i^H[n]\right)\omega_i[n]\right]^*,$$

$$\frac{\partial f_2}{\partial\sqrt{\mathbf{p}_i[n]}} = -\sum_{i=1}^{K}\sum_{n=1}^{N}|\omega_i[n]|^2\left[\left(\mathbf{d}_i^H[n] + \theta^H[n]\mathbf{R}_i[n]\right)\left(\mathbf{d}_i[n] + \theta[n]\mathbf{R}_i^H[n]\right)\right]^T\left(\sqrt{\mathbf{p}_i[n]}\right)^*,$$

$$\frac{\partial f_3}{\partial\sqrt{\mathbf{p}_i[n]}} = -\beta\sum_{i=1}^{K}\sum_{n=1}^{N}\left(\sqrt{\mathbf{p}_i[n]}\right)^*.$$

Solve the equation $\partial f_{A2}\big/\partial\sqrt{\mathbf{p}_i[n]} = 0$ to obtain the solutions of parameters $\mathbf{p}_i[n]$ as follows

$$\mathbf{p}_i[n] = (1+\ddot{\mu}_i[n])\ddot{\omega}_i^2[n]\ddot{\mathbf{A}}_j[n]\left(\ddot{\mathbf{B}}_i[n]\,\ddot{\mathbf{B}}_i^H[n]\right)\ddot{\mathbf{A}}_j^H[n], \tag{23}$$

where $\ddot{\mathbf{p}}$, $\ddot{\boldsymbol{\mu}}$, $\ddot{\boldsymbol{\omega}}$ represent the updated value of the parameters $\mathbf{p}$, $\boldsymbol{\mu}$, $\boldsymbol{\omega}$, respectively, $\ddot{\mathbf{A}}_j[n]$ and $\ddot{\mathbf{B}}_j[n]$ in formula (23) are expressed as

$$\ddot{\mathbf{A}}_j[n] = \left(\beta\mathbf{E}_{\mathbf{M}} + \sum_{j=1}^{K}|\ddot{\omega}_j[n]|^2\ddot{\mathbf{B}}_j[n]\,\ddot{\mathbf{B}}_j^H[n]\right)^{-1}, \tag{24}$$

$$\ddot{\mathbf{B}}_j[n] = \mathbf{d}_j[n] + \mathbf{R}_j^H[n]\,\ddot{\boldsymbol{\theta}}[n], \tag{25}$$

where $\mathbf{E}_M$ is the M-order unit matrix.

Algorithm 2 Prox-linear algorithm for solving problem PA

1: Initialize $\{\mathbf{P}^0, \boldsymbol{\mu}^0, \boldsymbol{\omega}^0\}$, and set iteration index $t^A = 0$.
2: **repeat**
3: Update $\boldsymbol{\omega}$ by (21).
4: Update $\mathbf{P}$ by (22).
5: Update $\boldsymbol{\mu}$ by (19).
6: Update $\boldsymbol{\omega}$ by (21).
7: Update $t_1^A = t_1^A + 1$.
8: **until** $t^A = t_{\max}^A$

B. IRS Reflecting Design

For IRS reflecting design, an IRM algorithm is proposed to design the phase shift. According to the updated results from the last subsection, auxiliary variables $\boldsymbol{X}$ and $\boldsymbol{\alpha}$ are introduced to transform the problem P(A) into a non-convex quadratic constrained quadratic programming (QCQP) problem. The new form P(B) for IRS phase shift optimization can be rewritten as

$$\begin{aligned} P(B):\ &\min_{\boldsymbol{\theta}} f_B(\boldsymbol{\theta}) \\ &s.t \quad |\theta_m| = 1, \forall m = 1, ..., M \end{aligned} , \tag{26}$$

where $f_B(\boldsymbol{\theta}) = \boldsymbol{\theta}^H \boldsymbol{X} \boldsymbol{\theta} - 2\mathrm{Re}\left\{\boldsymbol{\theta}^H \boldsymbol{\alpha}\right\}$, $\boldsymbol{X}$ and $\boldsymbol{\alpha}$ are

$$\boldsymbol{X} = \sum_{i=1}^{K} |\ddot{\omega}_i|^2 \sum_{j=1}^{K} \ddot{\boldsymbol{\lambda}}_{j,i} \ddot{\boldsymbol{\lambda}}_{j,i}^H \ , \tag{27}$$

$$\boldsymbol{\alpha} = \sum_{i=1}^{K} \left(\sqrt{(1 + \ddot{\mu}_i[n])} \ddot{\omega}_i^* \ddot{\boldsymbol{\lambda}}_{i,i} - |\omega_i|^2 \sum_{j=1}^{K} \ddot{\tau}_{j,i}^* \ddot{\boldsymbol{\lambda}}_{j,i} \right) \ , \tag{28}$$

where $\ddot{\boldsymbol{\lambda}}_{j,i} = \mathbf{R}_i \sqrt{\ddot{\mathbf{p}}_j}$ and $\ddot{\tau}_{j,i} = \mathbf{d}_i^H \sqrt{\ddot{\mathbf{p}}_j}$.

At first, a parameter $t^2 = 1$ is introduced to transform the non-homogeneous function f_B into homogeneous form

$$\boldsymbol{\theta}^H \boldsymbol{X} \boldsymbol{\theta} - 2\mathrm{Re}\left\{\boldsymbol{\theta}^H \boldsymbol{\alpha}\right\} = \begin{bmatrix} \boldsymbol{\theta} \\ t \end{bmatrix}^H \begin{bmatrix} \boldsymbol{X} & -\boldsymbol{\alpha} \\ -\boldsymbol{\alpha}^H & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ t \end{bmatrix} = \tilde{\boldsymbol{\theta}}^H \boldsymbol{X}' \tilde{\boldsymbol{\theta}} \ , \tag{29}$$

where $\boldsymbol{X}' = \begin{bmatrix} \boldsymbol{X} & -\boldsymbol{\alpha} \\ -\boldsymbol{\alpha}^H & 0 \end{bmatrix}$, $\tilde{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{\theta} \\ t \end{bmatrix}$.

Therefore, only the equivalent homogeneous formula in (29) needs to be solved. Then, the following method for non-convex QCQP problems only focus on the positive semidefinite matrix of rank one in homogeneous QCQP.

Finding a rank one matrix is computationally complex, especially for large-scale QCQP. This paper first introduces a continuous function, and approximates the rank function of a matrix with a given precision by designing appropriate parameters. Then the Rank Minimization Problems (RMP) is transformed into a rank-constrained optimization problem. An IRM algorithm [38] is employed to asymptotically approximate the constrained rank.

At present, in order to achieve computational efficiency, the existing methods usually bring in approximate functions instead of rank functions to establish the relaxation optimization problem, but this will sacrifice the optimality. In the following, an alternative method is proposed to re-express RMP as a rank-constrained optimization problem.

Then a rank-one positive semidefinite matrix $\boldsymbol{\Phi} = \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H$ is introduced, and the non-convex QCQP problem is transformed into the semidefinite programming problem P(B1) as follows

$$\begin{aligned} P(B1):\ & \min_{\Phi} \ \langle \boldsymbol{\Phi}, \boldsymbol{X} \rangle \\ s.t \quad & \boldsymbol{\Phi}_{m,m} = 1, \forall m = 1, ..., M \ , \\ & \boldsymbol{\Phi} \succeq 0 \end{aligned} \tag{30}$$

where $\langle \boldsymbol{\Phi}, \boldsymbol{X} \rangle = trace(\boldsymbol{\Phi}^T \boldsymbol{X})$. However, the rank one constraint on $\boldsymbol{\Phi}$ is nonlinear, and it is computationally complicated to find an accurate rank one solution. One of the present methods is to relax the rank 1 constraint through the semi-definite constraint, denoted as $\boldsymbol{\Phi} \succeq \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^H$. Compared with the linearized relaxation technique method, the semidefinite relaxation method usually produces a stricter lower bound of the optimal value. Nevertheless, the relaxation method fails to produce optimal solutions for unknown variables, and in most cases does not even produce infeasible solutions. Thus this paper introduce a new scheme for RMPs in the following.

Since the dimensionality of the target space can be represented by the trace based on the projection matrix as follows

$$\begin{aligned} & P(\boldsymbol{\Phi}) = \boldsymbol{\Phi}(\boldsymbol{\Phi}^H \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^H \\ & trace(P(\boldsymbol{\Phi})) = rank(\boldsymbol{\Phi}) \end{aligned} \ , \tag{31}$$

where $\boldsymbol{\Phi}^H \boldsymbol{\Phi}$ is non-singular, the trace of the projection matrix is equal to the rank of $\boldsymbol{\Phi}$. Consider the singular case, a auxiliary regularization parameter o is employed, and rewrite $P(\boldsymbol{\Phi})$ as

$$P_o(\boldsymbol{\Phi}) = \boldsymbol{\Phi}(\boldsymbol{\Phi}^H \boldsymbol{\Phi} + oI_n)^{-1} \boldsymbol{\Phi}^H \ , \tag{32}$$

where $trace(P_o(\mathbf{\Phi}))$ can be infinitely close to $rank(\mathbf{\Phi})$ and $trace(P_o(\mathbf{\Phi}))$ for $\mathbf{\Phi}$ is continuously differentiable. Thus the problem P(B1) can be rewritten as

$$\begin{aligned} P(B2): \min_{\mathbf{\Phi},\mathbf{Y}} & \operatorname{trace}(\mathbf{Y}) \\ s.t \quad & \mathbf{\Phi}_{m,m}=1, \forall m=1,...,M \\ & \mathbf{Y} \succeq \mathbf{\Phi}(\mathbf{\Phi}^H\mathbf{\Phi}+o\mathbf{E}_m)^{-1}\mathbf{\Phi}^H \end{aligned} , \tag{33}$$

where $\mathbf{Y}\in\mathbb{S}^m$ is the variable of the relaxed symmetric matrix. Introduce a new variable $\mathbf{Z}=\mathbf{\Phi}^H\mathbf{\Phi}$, $\mathbf{Z}\in\mathbb{S}^m$, and exploit Schur complement to transform the nonlinear matrix inequality in P(B2) into linear, which is converted to P(B3)

$$\begin{aligned} P(B3): \min_{\mathbf{\Phi},\mathbf{Y}} & \operatorname{trace}(\mathbf{Y}) \\ s.t \quad & \mathbf{\Phi}_{m,m}=1, \forall m=1,...,M \\ & \begin{bmatrix} \mathbf{Y} & \mathbf{\Phi} \\ \mathbf{\Phi}^H & \mathbf{Z} \end{bmatrix} \succeq 0 \end{aligned} . \tag{34}$$

In the new formula, matrix inversion is needless. Meanwhile, the new transformation is eliminated the regularization parameter o. Since $\mathbf{Z}=\mathbf{\Phi}^H\mathbf{\Phi}$ is non-convex, it is necessary to transform the RMP into a rank-constrained optimization problem.

It is proved by [38] that when $\mathbf{Z}\in\mathbb{S}^m$, $\mathbf{Z}=\mathbf{\Phi}^H\mathbf{\Phi}$ is equivalent to $rank(\begin{bmatrix} \mathbf{E}_m & \mathbf{\Phi} \\ \mathbf{\Phi}^H & \mathbf{Z} \end{bmatrix}) \le m$ and $\begin{bmatrix} \mathbf{E}_m & \mathbf{\Phi} \\ \mathbf{\Phi}^H & \mathbf{Z} \end{bmatrix} \succeq 0$. Therefore, transform RMP into rank-constrained optimization problem as follows

$$\begin{aligned} P(B4): \min_{\mathbf{\Phi},\mathbf{Y}} & \operatorname{trace}(\mathbf{Y}) \\ s.t \quad & \mathbf{\Phi}_{m,m}=1, \forall m=1,...,M \\ & \begin{bmatrix} \mathbf{Y} & \mathbf{\Phi} \\ \mathbf{\Phi}^H & \mathbf{Z} \end{bmatrix} \succeq 0 \\ & rank(\begin{bmatrix} \mathbf{E}_m & \mathbf{\Phi} \\ \mathbf{\Phi}^H & \mathbf{Z} \end{bmatrix}) \le m \end{aligned} . \tag{35}$$

The problem in (35) can be solved iteratively, and at rth iteration, the subproblem is expressed

as

$$\begin{aligned} P(B5) \ : \min_{\mathbf{\Phi}_r, \mathbf{Y}_r, \mathbf{Z}_r, e_r} \ & \operatorname{trace}(\mathbf{Y}_r) + \varpi^r e_r \\ s.t \quad & \mathbf{\Phi}_{m,m} = 1, \forall m = 1, ..., M \\ & \begin{bmatrix} \mathbf{Y}_r & \mathbf{\Phi}_r \\ \mathbf{\Phi}_r^H & \mathbf{Z}_r \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} \mathbf{E}_m & \mathbf{\Phi}_r \\ \mathbf{\Phi}_r^H & \mathbf{Z}_r \end{bmatrix} \succeq 0 \\ & e_r \mathbf{E}_m - \mathbf{V}_{r-1}^H \begin{bmatrix} \mathbf{E}_m & \mathbf{\Phi}_r \\ \mathbf{\Phi}_r^H & \mathbf{Z}_r \end{bmatrix} \mathbf{V}_{r-1} \succeq 0 \end{aligned}, \tag{36}$$

where $\varpi > 1$ is the weighting coefficient of e_r, $\mathbf{V}_{r-1}$ is the eigenvector of the m smallest eigenvalues corresponding to the solution of $\begin{bmatrix} \mathbf{E}_m & \mathbf{\Phi}_{r-1} \\ \mathbf{\Phi}_{r-1}^H & \mathbf{Z}_{r-1} \end{bmatrix}$ at the $(r-1)$th iteration.

In addition, in the first iteration $r = 1$, $\mathbf{V}_0$ needs to be initialized. Since the tracking heuristic method is easy to implement, the method is adopted to obtain the trace heuristic method of RMP represented in the initial input of $\mathbf{V}_0$ [39].

It can be seen that the solution at the convergence point satisfies the rank one constraint on $\mathbf{\Phi}$ and the other constraints proposed in the equivalent QCQP problem. On the basis of the Karush-Kuhn-Tucker condition, at least one locally optimal linear convergence of the proposed IRM method can be obtained. The proposed algorithm based on IRM is shown in Algorithm 3.

C. UAV Trajectory Optimization

According to the updated UAV transmit power $\mathbf{P}$ and IRS phase shift $\boldsymbol{\theta}$, the joint optimization problem is expressed as a subproblem P(C) about UAV trajectory optimization. The subproblem P(C) is represented as follows

$$\begin{aligned} P(C) : \ & \max_{\mathbf{Q}} f(\mathbf{Q}) = \sum_{i=1}^{K} \sum_{n=1}^{N} \log(1 + SINR_i[n]) \\ s.t \quad & C3, C4, C5. \end{aligned}. \tag{37}$$

For the subproblem P(C), applying for an enhanced reinforcement learning algorithm to update the UAV trajectory.

The horizontal target space of UAV trajectory is divided into grids with $(V_{\max}T/N)*(V_{\max}T/N)$, and different grids are converted into state space according to coordinates.

Algorithm 3 IRM-based algorithm for solving problem PB

Initialize $\boldsymbol{\theta}^0$.
repeat
  Update $\boldsymbol{X}$ and $\boldsymbol{\alpha}$ by (27) and (28) respectively.
  repeat
    Initialize set $r = 0$ and obtain $\mathbf{V}_0$.
    $r = r + 1$.
    while $e^r > o$ **do**
      solve problem (36) and obtain $\mathbf{\Phi}_r, \mathbf{Y}_r, \mathbf{Z}_r, e_r$.
      Update $\mathbf{V}_r$ from $\begin{bmatrix} \mathbf{E}_m & \mathbf{\Phi}_r \\ \mathbf{\Phi}_r^H & \mathbf{Z}_r \end{bmatrix}$
      $r = r + 1$.
    end while
  until find Φ.
  Update $\boldsymbol{\theta}$.
  Update $t^B = t^B + 1$.
until $t^B = t^B_{\max}$

In general, UAV can have multiple flight actions at a certain time slot in the permitted flight service area. For the convenience of training, the action space of the UAV is approximately divided into discrete action sets in multiple directions. In this paper, the 360-degree plane space is divided into 8 action/flight directions according to the 45-degree turning angle, add the action of the UAV that remain unchanged, the total action space of the UAV is 9.

Set the sum-rate of all the UE in a time slot as the reward function

$$R[n] = -c + \sum_{i=1}^{K} \log(1 + SINR_i[n]) + Penalty \ , \tag{38}$$

where c represents a constant, used to guide the UAV to the destination. Specifically, c serves as a reward for the punishment agent to take additional steps. Using this reward, it can motivate the agent to complete the task as soon as possible. $Penalty$ denotes the penalty coefficient, which the agent will be punished when it takes action in the no-fly zone [40].

The value function iterative updating formula of trajectory optimization by reinforcement learning is as follows

$$Q_{n+1}(s[n], a[n]) = (1-u)Q_n(s[n], a[n]) + u\left[R[n] + v \max_{a \in A} Q_n(s[n+1], a[n])\right] \ , \tag{39}$$

where $Q_n(s[n], a[n])$ is the value function, u is the learning rate factor, v is the discount factor, $a[n]$, $s[n]$ represent the action taken by agent and the state in the time slot n.

The artificial potential field method is applied to initialize the state space, so that the closer to the target position, the larger the state value is. This will guide the agent to move towards the target location and reduce a mass of invalid iterations caused by environmental exploration in the initial stage of the algorithm. The artificial potential field function is

$$U_{\text{att}} = \frac{\Delta}{|d| + \eta} \ , \tag{40}$$

where Δ is a scale factor greater than 0, used to adjust the size of gravity. $|d|$ is the distance between the current location and destination, and η denotes a normal number to prevent the gravitational value at the target point from appearing infinity. In the constructed artificial potential field, the whole potential field shows a monotonically increasing trend from the start point to the destination, and the destination has the maximum potential energy but is not infinite.

Consequently, during the initialization process the formula (39) is rewritten as

$$Q_n(s[n], a[n]) = R[n] + v \sum P(s[n+1] \,|s[n], a[n])V\left(s[n+1]\right) \ , \tag{41}$$

where $V\left(s[n+1]\right) = U_{\text{att}}$, $P(s_{n+1} \,|s_n, a_n)$ is the probability of transferring to the state $s[n+1]$, when the current state $s[n]$ and action $a[n]$ are determined, and $V\left(s[n+1]\right)$ is the state value function of the next state.

The balance between exploration and utilization is critical to reinforcement learning [41]. To a certain extent, ε- greedy strategy balance the exploration and utilization. However, the agent randomly selects actions in the action set with the probability of ε each time, and bad actions are also selected with the same probability. Therefore, the convergence speed of the whole process will be slow. Even if it converges at the end, the result will fluctuate due to the random selection of actions with the probability of ε. In response to this problem, an improved ε-greedy strategy is proposed to dynamically adjust the greedy factor as follows

$$\varepsilon = \begin{cases} \varepsilon_{\max}, & \text{if } \tanh(std_n/T) > \varepsilon_{\max} \\ \varepsilon_{\min}, & \text{if } \tanh(std_n/T) < \varepsilon_{\min} \\ \tanh(std_n/T), & \text{else} \end{cases} \ , \tag{42}$$

where $\tanh(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}}$. When $t>0$, $\tanh(t) \in [0, 1]$. std_n denotes the standard deviation of steps for n consequent iterations. T is the coefficient which the larger the T, the smaller the randomness. $\varepsilon_{\max}$ and $\varepsilon_{\min}$ is the maximum value and minimum value of the exploration rate respectively.

Algorithm 4 Enhanced RL algorithm for solving problem PC

1: Set Action space A, State space S, Maximum iterations $t_{\max}^{C1}$, Maximum trials $t_{\max}^{C2}$, learning parameters v, u.
2: Initialize S by (36).
3: **repeat**
4: **repeat**
5: Update strategy by (42).
6: Choose action from the available actions in the current state $s[n]$.
7: Take action $a[n]$, and update $s[n+1]$ by (39).
8: Update $t^{C1} = t^{C1} + 1$.
9: **until** Convergence or $t^{C1} = t_{\max}^{C1}$.
10: Update $t^{C2} = t^{C2} + 1$.
11: **until** Convergence or $t^{C2} = t_{\max}^{C2}$.

Due to the algorithm does not converge in the initial stage and the stdn is large, the agent randomly chooses actions with the probability of $\varepsilon_{\max}$. As the algorithm progresses, the stdn decreases so that ε takes a value in the range of $(\varepsilon_{\min}, \varepsilon_{\max})$. The greater the stdn, the greater the difference in the number of steps between iterations, the more the environment needs to be explored, and the greater the value of ε. When stdn is small, it indicates that the algorithm tends to converge, and ε is stable at $\varepsilon_{\min}$. It can be seen from the above analysis that the dynamic adjustment strategy of the greed factor designed by this algorithm enables the environment to be explored with a greater probability in the early stage. As the algorithm progresses, it gradually tends to be utilization, which can better balance the exploration and utilization [42].

In addition, apply for Eligibility Traces (ET) mechanism to the constructed reinforcement learning framework, the learning speed of the proposed algorithm can be significantly improved. Specifically, the algorithm can record the number of times the state is accessed, and when the state value function is updated at the previous moment, the previous state value function can also be updated [43]. The incremental ET is calculated as follows

$$e_n(s) = \begin{cases} \lambda v e_{n-1}(s), & \text{if } \mathrm{s} \neq s[n] \\ 1, & \text{if } \mathrm{s} = s[n] \end{cases}. \tag{43}$$

Finally, after several explorations, the value function gradually approaches to the optimal value function, and finally the trajectory optimization of the UAV is realized. The whole process is updated iteratively to achieve convergence conditions. The proposed algorithm based on enhanced reinforcement learning is shown in Algorithm 4.

D. Complexity Analysis

In this subsection, the computational complexity of the joint optimization algorithm is given in the following. For Algorithm 2, the complexity to update $\boldsymbol{\omega}$, $\boldsymbol{\mu}$, and $\mathbf{P}$ are $\mathcal{O}(KM)$, $\mathcal{O}(KM)$, and $\mathcal{O}(K)$, respectively. Thus the asymptotic time complexity can be expressed as $\mathcal{O}(2KM+K)$. For Algorithm 3, the sub-problem based on the SDP can be solved by the interior-point method, the order of complexity for a SDP problem with m SDP constraints which includes an $n \times n$ positive semi-definite matrix is given by $\mathcal{O}\left(\sqrt{n}\log(1/o)\left(mn^3+m^2n^2+m^3\right)\right)$, where $o>0$ is the solution accuracy. For problem $P(B5)$, with $n=M+1$, $m=1$, the approximate computational complexity for solving $P(B5)$ can be written as $\mathcal{O}\left(\log(1/o)\left(2(M+1)^{3.5}+(M+1)^{2.5}\right)\right)$. As for Algorithm 4, the complexity includes two parts: computational complexity $\mathcal{O}\left(t_{\max}^{C1}\right)$ and training complexity $\mathcal{O}\left(t_{\max}^{C1}t_{\max}^{C2}\right)$. The asymptotic time complexity can be expressed as $\mathcal{O}\left(t_{\max}^{C1}+t_{\max}^{C1}t_{\max}^{C2}\right)$. Therefore, the complexity of the overall joint optimization algorithm is $\mathcal{O}\left(t_{\max}^{A}\left(2KM+K+t_{\max}^{B}\left(\log(1/o)\left(2(M+1)^{3.5}+(M+1)^{2.5}\right)\right)\right)+t_{\max}^{C1}+t_{\max}^{C1}t_{\max}^{C2}\right)$.

IV. Simulation Results

This section gives simulation analysis to demonstrate the validity of the proposed scheme. The simulation parameters are set in the following. The fixed location of IRS is (100,0,20). The number of IRS reflection elements can be flexibly set along with the needs of each experiment. The referenced channel power gain is set to $\rho^0=-50dB$, and the noise power $\sigma^2=-80dB$. The path loss index of the UAV-UE link is set to 2.5, and the corresponding Rician factor is 10 dB. The path loss index and Rician factor of the IRS-UE link are 2.2 and 10 dB, respectively. The maximum transmission power is set to 10dB and the height of the UAV is 30m. Within range of $[0,2\pi]$, the phase shift of the reflection unit is randomly and uniformly generated. Finally, this paper considers K = 4 users, who are randomly and evenly distributed in an area of $200\times200m^2$. The simulation parameters for applying reinforcement learning for trajectory optimization are as follows: the learning rate u is 0.02, the discount factor v is 0.9, the maximum number of iterations is 20000, the scale factor Δ is 0.6, and the constant η is 1. The greedy factor dynamically adjusts the strategy parameters as follows, $\varepsilon_{\max}$=0.5, $\varepsilon_{\min}$=0.01. The no-fly zone simulated in this paper refers to any place outside the rectangular area where users are evenly distributed.

In Fig. 2, the convergence of the proposed scheme versus different UAV maximum transmit

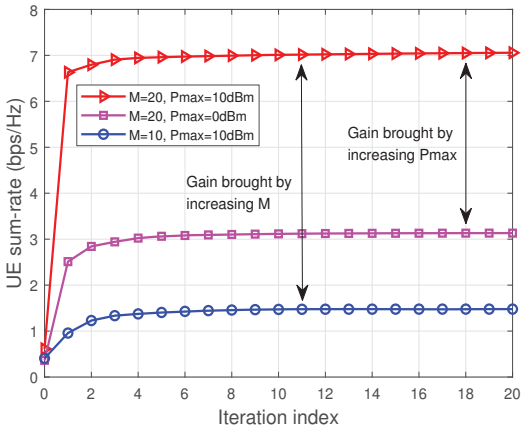


Fig. 2. The convergence of the proposed scheme for different cases.

power, Pmax and the reflecting elements number of IRS, M is presented. This paper considers the following three schemes: 1) M = 20, $P_{\max}$ = 10 dBm; 2) M = 20, $P_{\max}$ = 0 dBm; 3)M = 10, $P_{\max}$ = 10 dBm. Uniformly, the proposed algorithm for the three cases converges with the increase of iterations. For these three schemes, the convergence times of the proposed algorithm in this paper is all approximately around 5 times. Moreover, it can be discovered that the sum-rate increase as the number of IRS increases, since the proposed phase shift design can realize the enhancement of passive beamforming gain. Specifically, compared with scheme3, scheme2 shows an increase in the number of IRS by 10, and the user's sum-rate gain has increased by 78%. Increasing the transmit power of UAV can also improve sum-rate of the system. Fig. 2 also presents that compared with scheme 1 and scheme 2, when Pmax is increased by 10dBm, the sum-rate gain is increased by 55%. From the above analysis, it can be seen that the system gain brought by increasing the number of IRS is significant compared with increasing the transmit power of UAV. Due to the convenience and cheapness of IRS deployment, the deployment of IRS has played a cost-saving effect to a large extent compared to the method of increasing system capacity by significantly increasing expensive transmit power resources.

Fig. 3 presents the changes in cumulative sum-rate under different methods verus different UAV flight times. Compared the proposed scheme with the initial trajectory scheme, it can be concluded that the trajectory optimization algorithm based on enhanced reinforcement learning in this paper elevates the system performance gain by 32%. The effectiveness and high performance

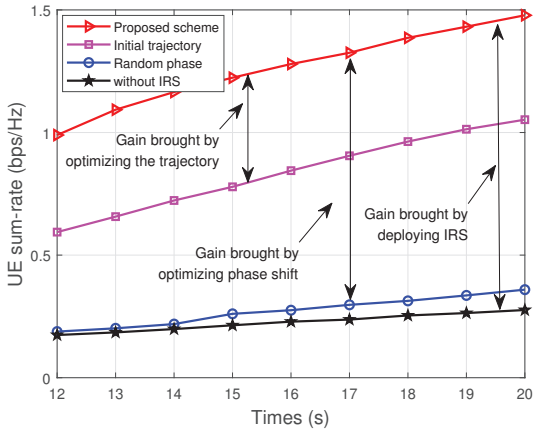


Fig. 3. The different methods versus different flight times of UAV.

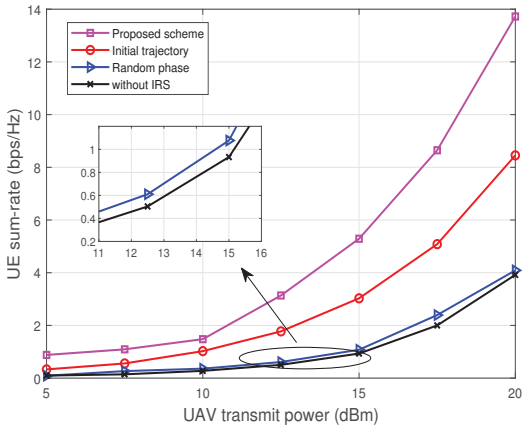


Fig. 4. The different methods versus the maximum transmit power of UAV.

gains of the proposed UAV trajectory optimization scheme are proved. Similarly, only change the phase shift optimization scheme in the joint optimization scheme, without changing the power control and trajectory optimization methods, and the performance gain of the system has increased by an astonishing 81%. It proves that our proposed phase shift optimization scheme is excellent. Finally, compared the proposed scheme with the benchmark scheme without the deployment of IRS, the performance gains are noteworthy. Last but not least, the performance gain of the random phase shift scheme is negligible compared to the proposed phase shift optimization scheme. This verifies the importance of IRS phase shift design.

Fig. 4 exhibits the relationship between cumulative sum-rate of different schemes versus

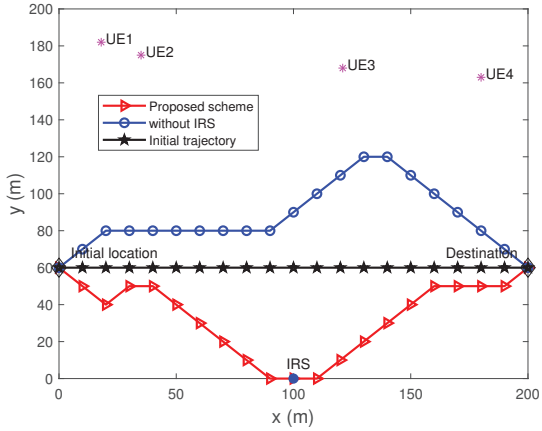


Fig. 5. The trajectory of UAV under different schemes.

different Pmax is explored. The results show that for all the schemes, the sum-rate of users improve with the growth of the maximum transmit power. It is also clear that deploying IRS in the system is significantly better than without IRS, especially if $P_{\max}$ is relatively large. According to the comparison between random phase shift scheme and the joint optimization scheme based on IRM algorithm, it can be concluded that the phase shift design proposed in this paper is remarkable to improve performance. The gain curves of the joint optimization scheme based on random phase shift and without IRS tend to be close in the whole process, even if the transmit power of the UAV continues to increase. Therefore, there are sufficient reasons to prove that if the phase shift optimization design is not carried out, the deployment of IRS will not be able to show its original huge performance gains.

Fig. 5 shows the trajectory of UAV under three different schemes. When the UAV communicates with $K = 4$ users, and there is no IRS in the environment, the UAV tends to fly to the place where the users are concentrated to reduce the path loss. As for the scenarios where IRS is deployed, the UAV tends to the nearby area where the IRS is deployed for keeping a high system gain. Due to the task of flying to the destination, the UAV slowly leave the vicinity of the IRS deployment area and fly towards the destination. Obviously, the deployment of IRS brings a significant increase in the sum-rate of users, IRS can effectively help any user in need, and its related performance gains even exceed the corresponding performance gains brought by UAV trajectory optimization, which can be seen in the Fig. 3.

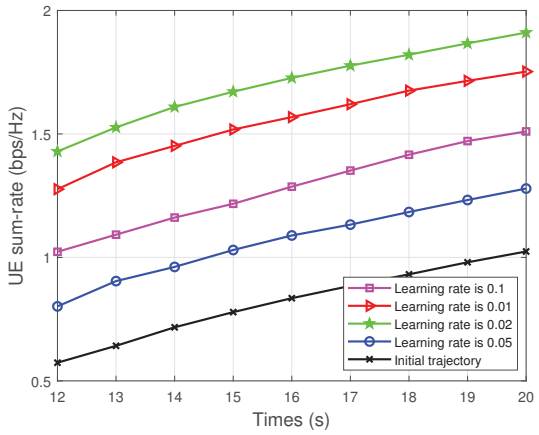
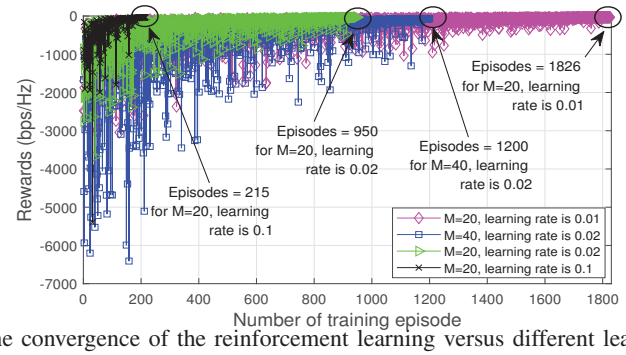


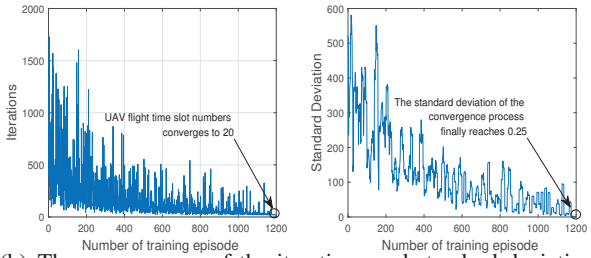
Fig. 6. The learning rate of reinforcement learning, the initial trajectory versus times.

The choice of network parameters determines the convergence speed and efficiency of learning usually. We take the learning rate of reinforcement learning network as an example to illustrate the importance. Fig. 6 shows a comparison of system gains brought by trajectory optimization under different learning rates. obviously, different learning rates play different roles in the performance of the reinforcement learning algorithm. It can be discovered that when the learning rate is set to 0.1, its reward gain is much lower than the learning rate of 0.02. This is because when the learning rate is too large, behavior oscillations will be occurred. In addition, if the learning rate is set too small, such as 0.01, it will take longer time to reach convergence, which can be seen in Fig. 7(a). Due to the concept of learning rate, it helps to achieve a compromise between training speed and the convergence. Apparently, the learning rate which is set to 0.02, the model learns the problem well. Therefore, we can choose an appropriate learning rate, which is neither too large nor too small, in our abundant experiments, the learning rate can be setting around 0.02.

Fig. 7 shows the convergence of the reward, iterations and standard deviation based on reinforcement learning. Fig. 7(a) first compares the reward convergence brought by the trajectory optimization with different learning rate. Apparently, the higher the learning rate, the faster the reward convergence speed. For instance, when the learning rate is 0.1, the algorithm convergence only needs 215 times. However, Fig. 6 exhibits that the higner the learning rate does not always lead to greater rewards. This is because setting the learning rate too large may cause the agent to explore the environment incompletely, and get sub-optimal results in the end. Similarly, if the



(a) The convergence of the reinforcement learning versus different learning rate.



(b) The convergence of the iterations and standard deviation.

Fig. 7. The convergence of rewards, iterations and standard deviation.

learning rate is set too small, such as 0.01, it will take a lot of time to converge the algorithm, for about 1826 times. In addition to the above experiments, Fig. 7(a) also compares the convergence of the agent's reward with different M. It can be seen from the figure that for the same learning rate of 0.02, the bigger the M, the more the number of convergence times the algorithm needs, which are 950 and 1200 respectively in the case of M=20 and M=40. At last, Fig. 7(b) shows the convergence of iterations and standard deviation in the process of UAV trajectory optimization when M=40. It can be seen that the UAV finally reached the convergence target of the specified step number 20 and standard deviation 0.25 at approximately 1200th training episodes.

Fig. 8 shows the impact of changing the number of users on the performance gain of the proposed scheme. Obviously, when the number of users increases, the sum rate brought by the proposed scheme increases significantly. It can be seen from the figure that in the case of the same number of IRS reflecting elements, the greater the maximum transmit power of the UAV, the greater the sum-rate of the system. In the case of the same transmit power, the increase in the number of IRS elements also leads to a larger performance gain. This is because that with the increase number of users enables the UAV to fly from the n-th slot to the $(n+1)$-th slot to serve more users, thereby improving the sum-rate of the system. This means that the proposed

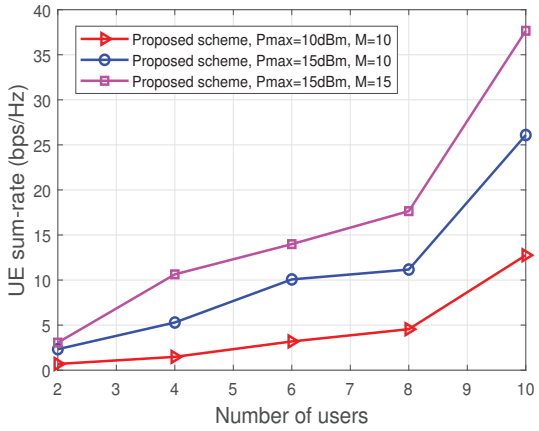


Fig. 8. The different methods versus the number of users.

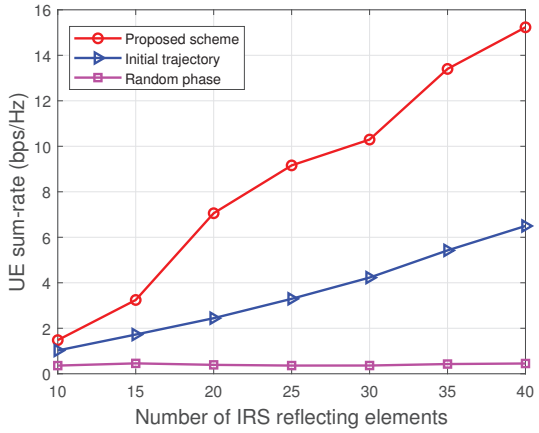


Fig. 9. The different methods versus the number of IRS reflecting elements.

algorithm is more suitable for scenarios with more users, and it is in this scenario that UAV make more economic sense.

Fig. 9 exhibits the relationship between the sum-rate of the users and the number of IRS reflecting elements. It can be seen that the performance gain of the proposed scheme increases with the number of IRS reflecting elements, because the more reflecting elements, the higher the passive beamforming gain the system can obtain. Furthermore, the IRS random phase shift scheme has very poor performance and only a small performance gain is obtained, while as M increases, both the proposed scheme and the initialization trajectory have significant performance gains. It can also be seen from the figure that under the condition of the same number of IRS

reflecting elements, the effect of the trajectory optimization scheme of the proposed algorithm is very obvious compared with the performance gain under the initialized trajectory.

V. Conclusion

This paper focuses on the transmit power, IRS phase shift design and UAV trajectory optimization of IRS empowered UAV communication network, and investigates the maximum user sum-rate. Specifically, the BCD method is used to decompose the proposed problem block by block, and two variables are fixed to optimize the third variable for alternate optimization. Firstly, the quadratic transformation and Lagrange dual transformation is used to transform the problem, and the approximate linear algorithm is adopted to optimize the UAV transmit power. Secondly, aiming at the QCQP problem of IRS phase shift, IRM method is brought up to optimize the phase shift. Finally, an enhanced reinforcement learning algorithm is employed to optimize UAV trajectory. In numerous simulation experiments, the proposed method is compared with the benchmark method to verify the superiority, and the user sum-rate is analyzed at different flight altitudes of UAV. The simulation results also show that the deployment of IRS has a remarkable performance on improving system gains. The reflective elements numbers and phase shift design of IRS can also greatly influence the sum-rate. Future works can be summarized as follows: 1) Study a robust optimization with channel estimation error. 2) Consider energy efficiency of UAV wireless networks based on uniform rectangular array for the IRS. 3) Employ DRL-based beamforming design for IRS.

References

[1] M. Mozaffari, W. Saad, M. Bennis, Y . Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart. 2019.

[2] M. D. Renzo, A. Zappone, M. Debbah, M. Alouini, C. Yuen, J. D. Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[3] X. Mu, Y. Liu, L. Guo, J. Lin, and H. V. Poor, "Intelligent reflecting surface enhanced Multi-UAV NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 3051-3066, Oct. 2021.

[4] M. Hua, L. Yang, Q. Wu, C. Pan, C. Li, and A. L. Swindlehurst, "UAV-assisted intelligent reflecting surface symbiotic radio system," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5769-5785, Sept. 2021.

[5] Z. Wei, Y. Cai, Z. Sun, D. W. K. Ng, J. Yuan, M. Zhou, and L. Sun, "Sum-rate maximization for IRS-assisted UAV OFDMA communication systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2530-2550, Apr. 2021.

[6] M. Gapeyenko, V. Petrov, D. Moltchanov, S. Andreev, N. Himayat, and Y. Koucheryavy, "Flexible and reliable UAV-assisted backhaul operation in 5G mmWave cellular networks,"*IEEE J. Sel. Areas Commun.*, vol. 36,no. 11, pp. 2486-2496, Nov. 2018.

[7] Y. Li, H. Zhang, K. Long, S. Choi, and A. Nallanathan, "Resource allocation for optimizing energy efficiency in NOMA-based fog UAV wireless networks," *IEEE Netw.*, vol. 34, no. 2, pp. 158-163, Mar/Apr. 2020.

[8] N. Zhao, F. Cheng, F. R. Yu, J. Tang, Y. Chen, G. Gui, and H. Sari, "Caching UAV Assisted Secure Transmission in Hyper-Dense Networks Based on Interference Alignment," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2281-2294, May. 2018.

[9] H. Zhang, J. Zhang, and K. Long, "Energy efficiency optimization for NOMA UAV network with imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2798-2809, Dec. 2020.

[10] X. Hu, K. Wong, K. Yang, and Z. Zheng, "UAV-assisted relaying and edge computing: Scheduling and trajectory optimization," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4738-4752, Oct. 2019.

[11] Q. Wu, Y. Zeng, and R. Zhang, "Joint trajectory and communication design for Multi-UAV enabled wireless networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2109-2121, Mar. 2018.

[12] J. Xu, Y. Zeng, and R. Zhang, "UAV-enabled wireless power transfer: trajectory design and energy optimization," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5092-5106, Aug. 2018.

[13] Y. Li, H. Zhang, K. Long, C. Jiang, and M. Guizani, "Joint resource allocation and trajectory optimization with QoS in UAV-based NOMA wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6343-6355, Oct. 2021.

[14] Y. Li, H. Zhang, and K. Long, "Joint resource, trajectory, and artificial noise optimization in secure driven 3D UAVs with NOMA and imperfect CSI," *IEEE J. Sel. Areas Commun.*, doi: 10.1109/JSAC.2021.3088623.

[15] X. Yuan, T. Yang, Y. Hu, J. Xu, and A. Schmeink, "Trajectory design for UAV-enabled multiuser wireless power transfer with nonlinear energy harvesting," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1105-1121. Feb. 2021.

[16] Z. Na, Y. Liu, J. Shi, C. Liu, and Z. Gao, "UAV-supported clustered NOMA for 6G-enabled Internet of Things: trajectory planning and resource allocation," *IEEE Internet of Things J.*, doi: 10.1109/JIOT.2020.3004432.

[17] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2329-2345, Apr. 2019.

[18] S. Hu, W. Ni, X. Wang, A. Jamalipour, and D. Ta, "Joint Optimization of Trajectory, Propulsion, and Thrust Powers for Covert UAV-on-UAV Video Tracking and Surveillance," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1959-1972, Jan. 2021.

[19] S. Hu, Q. Wu, and X. Wang, "Energy Management and Trajectory Optimization for UAV-Enabled Legitimate Monitoring Systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 142-155, Jan. 2021.

[20] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. Di Renzo, and M. Debbah, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118-125, Oct. 2020.

[21] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, Z. Qin, M. D. Renzo, and N. Al-Dhahir, "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Commun. Surv. Tut.*, vol. 23, no. 3, pp. 1546-1577, 3rd Quart. 2021.

[22] H. Xie, J. Xu, and Y. F. Liu, "Max-min fairness in IRS-aided multi-cell MISO systems with joint transmit and reflective beamforming," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1379-1393, Feb. 2021.

[23] Q. Pan, J. Wu, X. Zheng, W. Yang, and J. Li, "Differential privacy and IRS empowered intelligent energy harvesting for 6G Internet of Things," *IEEE Internet of Things J.*, doi: 10.1109/JIOT.2021.3104833.

[24] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394-5409, Nov. 2019.

[25] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157-4170, Aug. 2019.

[26] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6884-6898, Oct. 2020.

[27] H. Guo, Y. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064-3076, May. 2020.

[28] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 375-388, Jan. 2021.

[29] T. Jiang, H. V. Cheng, and W. Yu, "Learning to reflect and to beamform for intelligent reflecting surface with implicit channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 1931-1945, Jul. 2021.

[30] A. Ranjha and G. Kaddoum, "URLLC facilitated by mobile UAV relay and RIS: A joint design of passive beamforming, blocklength, and UAV positioning," *IEEE Internet of Things J.*, vol. 8, no. 6, pp. 4618-4627, Mar. 2021.

[31] A. Khalili, E. M. Monfared, S. Zargari, M. R. Javan, N. Mokari, and E. A. Jorswieck, "Resource management for transmit power minimization in UAV-assisted RIS HetNets supported by dual connectivity," *IEEE Trans. Wireless Commun.*, doi: 10.1109/TWC.2021.3107306.

[32] S. Li, B. Duo, M. Di Renzo, M. Tao, and X. Yuan, "Robust secure UAV communications with the aid of reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, doi: 10.1109/TWC.2021.3073746.

[33] X. Pang, M. Sheng, N. Zhao, J. Tang, D. Niyato, and K. -K. Wong, "When UAV Meets IRS: Expanding Air-Ground Networks via Passive Reflection," *IEEE Wireless Commun.*, vol. 28, no. 5, pp. 164-170, Oct. 2021.

[34] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2042-2055, Jul. 2021.

[35] R. Ikeshita, T. Nakatani, and S. Araki, "Block coordinate descent algorithms for auxiliary-function-based independent vector extraction," *IEEE Trans. Signal Process.*, vol. 69, pp. 3252-3267, 2021.

[36] K. Shen and W. Yu, "Fractional programming for communication systems—part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616-2630, May. 2018.

[37] K. Shen and W. Yu, "Fractional programming for communication systems—Part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631-2644, Mar. 2018

[38] C. Sun, N. Kingry, and R. Dai, "A unified formulation and nonconvex optimization method for mixed-type decision-making of robotic systems," *IEEE Trans. Robot.*, vol. 37, no. 3, pp. 831-846, Jun. 2021.

[39] W. Zeng and H. C. So, "Outlier-robust matrix completion via ℓ_p -minimization," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1125-1140, Mar. 2018.

[40] T. Shafique, H. Tabassum, and E. Hossain, "Optimization of wireless relaying with flexible UAV-borne reflecting surfaces," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 309-325, Jan. 2021.

[41] S. Hu, X. Chen, W. Ni, E. Hossain, and X. Wang, "Distributed Machine Learning for Wireless Communication Networks: Techniques, Architectures, and Applications," *IEEE Commun. Surv. Tut.*, vol. 23, no. 3, pp. 1458-1493, 3rd Quart. 2021.

[42] H. Zhang, N. Yang, W. Huangfu, K. Long, and V. C. M. Leung, "Power control based on deep reinforcement learning for spectrum sharing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4209-4219, Jun. 2020.

[43] Z. Zhang, R. Wang, F. R. Yu, F. Fu, and Q. Yan, "QoS Aware Transcoding for Live Streaming in Edge-Clouds Aided HetNets: An Enhanced Actor-Critic Approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11295-11308, Nov. 2019.