Zhang *et al.*
EURASIP Journal on Advances in Signal Processing (2023) 2023:111
https://doi.org/10.1186/s13634-023-01076-1

EURASIP Journal on Advances in Signal Processing

RESEARCH Open Access

Throughput maximization for irregular reconfigurable intelligent surface assisted NOMA systems

Weilin Zhang[1], Lingyi Wang[1], Hangtao Mao[1], Zi Wang[1] and Wei Wu[1*]

*Correspondence: weiwu@njupt.edu.cn

[1] School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Xianlin Street, Nanjing 210023, Jiangsu, China

Abstract

Reconfigurable intelligent surface (RIS) is an emerging technology to improve the spectral efficiency of wireless communication systems. However, the high complexity of beam design and the non-negligible overhead associated with RIS limit the number of elements that can be deployed in practice. In this paper, we investigate the downlink communications of irregularly deployed intelligent reflecting surfaces that assist non-orthogonal multiple access (NOMA) systems. To address this challenge, we propose a novel four-step resource allocation algorithm. Specifically, we first obtain a sub-optimal solution for the sparse deployment of RIS elements using a Simulated Annealing Algorithm. We then solve the power allocation problem by employing an integer optimization algorithm that continuously iterates the immobile point. To simplify and optimize the reflection coefficient matrix, we propose a construction inequality algorithm. Finally, we optimize the channel assignment using a genetic algorithm. The simulation results demonstrate that the proposed irregular RIS-assisted NOMA system outperforms the traditional RIS-assisted orthogonal multiple access system, with a maximum throughput increase of approximately 30%.

Keywords: Irregular intelligent reflecting surface, Non-orthogonal multiple access, Throughput maximization, Resource allocation, Low-order iterative optimization algorithm.

1 Introduction

At present, the development of 6 G technology is in full swing. 6 G is expected to bring an essential economic contribution [1], and the 6 G era is also considered to be a hyper-collaborative society in which various research findings in different fields will converge to open up a whole new world [2]. In the face of the needs of 6 G technology development, the Internet of Things (IoT) demands large connectivity, access to the large-scale device, and a shortage of spectrum resources, a number of technologies such as RIS and NOMA have begun to receive widespread attention.

RIS has the ability to reflect electromagnetic waves to expand the coverage of the base station (BS) and tune channels by adjusting the phase shift of each element, so it can improve the throughput of the system with limited resources and help alleviate the

problem of scarcity of spectrum resources. It can also reduce interference between users, improve the communication quality of edge users and guarantee the security of communication to a certain extent. At the same time, it improves the coverage of enhanced wireless signals, which has a strong advantage under perfect channel state information [3].

However, in practical systems, the performance of the traditional RIS is highly dependent on the number of RIS elements [4]. And the increase in the number of RIS elements not only results in very high channel state information acquisition costs but also leads to signal fading, which limits the capacity of RIS-assisted communication systems. Especially with a low Signal to Noise Ratio (SNR), the high number of RIS elements is causing significant performance losses [5]. Conversely, irregular RIS is irregularly configured with a certain number of RIS elements on an extended surface, which gives RIS elements additional freedom and increases the power and system capacity to receive signals. In [6], irregular RIS has also been shown to improve the performance of the MIMO model by up to 21%.

As the number of users accessing a system increases, the problem of interference becomes more severe, leading to a decrease in system throughput. To enhance the system's capacity, NOMA is employed to mitigate interference between users. NOMA enables multiple users to share the same time slot, frequency blocks, and channel code, thereby significantly increasing the system's throughput and spectral efficiency. Additionally, the successive interference cancelation (SIC) technique of NOMA effectively improves the SNR and reception reliability, which is crucial for information transmission. In [7], RIS-NOMA was shown to serve more users in each orthogonal spatial direction. For a given number of users and predefined transmit power, the total rate of NOMA is better than that of OMA [8]. Furthermore, a new algorithm has been proposed in [9] to suppress the error propagation of SIC.

There have been many articles on traditional RIS-assisted NOMA systems. Specifically, in [10], some methods for the maximum throughput of RIS-assisted NOMA networks are presented. In [11], the case of multiple clusters is considered and the minimum transmission power is obtained by a fixed point method. In [12], the minimum transmission power of RIS-assisted SWIPT NOMA networks is investigated by using a joint optimization approach.

The above articles consider the improvement of system performance when NOMA is combined with traditional RIS, but the work on system throughput when NOMA is combined with irregular RIS has not been carried out yet. Inspired by the irregular RIS-assisted multi-user multiple-input multiple-output (MIMO) communication system [6]. Similar to the stringent resource limitations discussed in [13] and the wireless channel characteristics outlined in [14], we investigate the scenario of downlink communication in the absence of direct links.

In this paper, we investigate the maximization of the throughput rate of irregular RIS-assisted NOMA systems to enhance system performance when the signal is not directly accessible. The primary contributions of this study can be summarized as follows:

(1) This study aims to maximize the system throughput of NOMA system assisted by irregular RIS and introduces a joint optimization problem. Compared to traditional

RIS-assisted NOMA systems, this system demonstrates significant performance improvements.

(2) To address the NP-hard problem, we propose a four-step joint optimization algorithm that sequentially employs sub-optimal algorithm, fixed-point iterative algorithm, constructive inequality algorithm, and Simulated Annealing Algorithm for stepwise optimization. By tackling this challenging non-convex optimization problem through stepwise optimization, we obtain optimized results while ensuring computational efficiency and real-time performance.

(3) We employ the Simulated Annealing Algorithm to solve the complex irregular RIS topology optimization problem. This algorithm optimizes the performance of the irregular RIS algorithm while maintaining low computational complexity, surpassing traditional methods. These results effectively demonstrate the effectiveness and efficiency of our proposed method.

The remainder of this paper is organized as follows: In Sect. 2, we construct the system model and define the optimization problem, as well as introduce the constraints. In Sect. 3, we propose a four-step joint optimization algorithm to solve the problem. Section 4 presents the simulation results, and Sect. 5 concludes the paper.

2 System model and problem formulation

2.1 System model

Similar to the scenario of multi-user grouping discussed in [15], we investigate an irregular RIS-assisted NOMA downlink wireless communication system, as depicted in Fig. 1. The system comprises a single-antenna BS and K users, each equipped

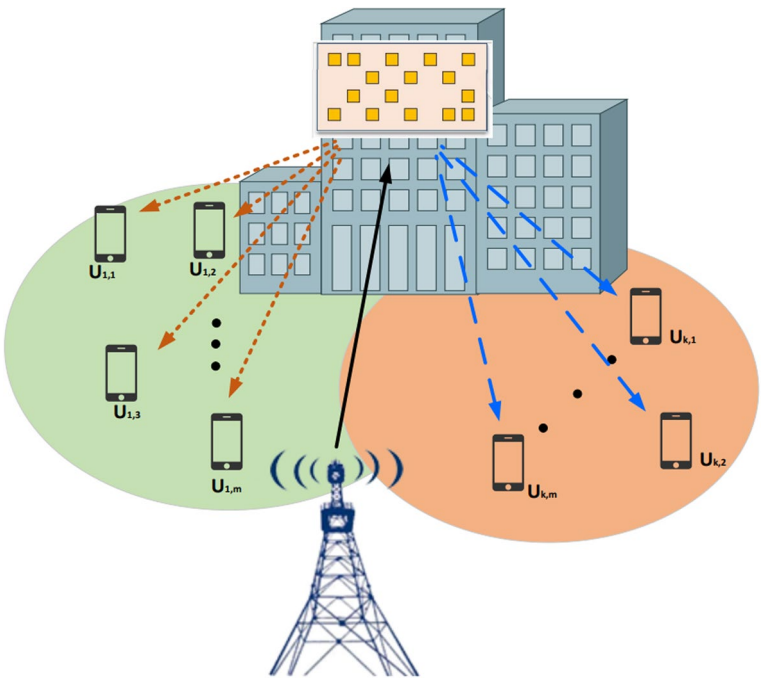

Fig. 1 Irregular RIS-assisted NOMA systems

with one antenna. The channels associated with the users are represented by the set $\boldsymbol{K} = \{1, 2, 3, \ldots, K\}$. The RIS, consisting of M reflecting elements sparsely distributed over M_{s} grid points, is assumed to have a spacing between adjacent grid points that is half of the signal wavelength [16]. The reflection coefficient matrix of the RIS is denoted as $\boldsymbol{\Theta} = \mathrm{diag}\left\{\lambda_1 e^{j\theta_1}, \lambda_2 e^{j\theta_2}, \lambda_3 e^{j\theta_3}, \ldots, \lambda_M e^{j\theta_M}\right\}$, where $\theta_m \in [0, 2\pi]$ and $\lambda_m \in [0, 1]$ represent the phase shift and amplitude of the mth reflecting element, respectively [17]. The selection matrix representing the topology of the RIS is denoted as $\boldsymbol{Z} = \mathrm{diag}\{\boldsymbol{z}\}$, where $\boldsymbol{z} = \left[z_1, z_2, z_3, \ldots, z_{M_{\mathrm{s}}}\right]^T$ and $z_i \in \{0,\ 1\}$, with "0" indicating selected and "1" indicating unselected. The number of ones in vector $\boldsymbol{z}$ is M, and the number of zeros is $M_{\mathrm{s}} - M$.

Due to the huge propagation loss [18], we just consider the signal reflected by the first time of the irregular RIS and no longer consider multiple reflections. Moreover, in order to better reflect the improvement of the signal by irregular RIS and the direct link is obscured in some practical cases, we only consider the reflected link and ignore the direct link.

The symbol to be transmitted on the nth channel is a superposition of signals from different users, can written as

$$x_n = \sum_{k=1}^{K} \delta_{n,k} \sqrt{p_{n,k}} s_{n,k} \tag{1}$$

where $\delta_{n,k}$ and $p_{n,k}$ represent the binary power control and power allocation variables, respectively. $s_{n,k}$ denotes the symbol to be transmitted by the kth user on the nth channel

User k receive the signal received over channel n can be written as

$$y_{n,k} = \left(\boldsymbol{g}_{\boldsymbol{n,k}}^{\boldsymbol{H}} \boldsymbol{Z} \boldsymbol{\Theta} \boldsymbol{f}_{\boldsymbol{n}}\right) x_n + n \tag{2}$$

where $\boldsymbol{g_{n,k}} \in \mathbb{C}^{M \times 1}$ and $\boldsymbol{f_n} \in \mathbb{C}^{M \times 1}$ denote the channel gain for channel n from the RIS to user k and from the BS to RIS, respectively. The channel gain for channel n between the BS and user k is ignored. n denotes white Gaussian noise (AWGN) with zero mean and σ^2 variance.

One of the key challenges in NOMA systems is determining the optimal SIC decoding order, which is typically based on the channel gain. However, in NOMA systems assisted by RIS, the combined channel gain can be adjusted by tuning the RIS reflection coefficients. This approach is also applicable to irregular RIS-assisted NOMA systems. Specifically, the decoding order $\pi_{n,k}$ for user k transmitting over channel n can be used to indicate the order in which the receiver should decode the signals, with $\pi_{n,k} = i$ denoting that the ith signal should be decoded first. The achievable capacity for user k on channel n can then be written as

$$R_{n,k} = \log_2 \left(1 + \frac{\delta_{n,k} p_{n,k} \left|\boldsymbol{g}_{\boldsymbol{n,k}}^{\boldsymbol{H}} \boldsymbol{Z} \boldsymbol{\theta} \boldsymbol{f}_{\boldsymbol{n}}\right|^2}{\left|\boldsymbol{g}_{\boldsymbol{n,k}}^{\boldsymbol{H}} \boldsymbol{Z} \boldsymbol{\theta} \boldsymbol{f}_{\boldsymbol{n}}\right|^2 P_{n,k} + \sigma^2}\right) \tag{3}$$

where $P_{n,k} = \sum_{\pi_{n,i} > \pi_{n,k}} \delta_{n,k} p_{n,k}$.

Assuming that $\pi_{n,k'} > \pi_{n,k}$, it is necessary to satisfy the SIC decoding condition $R_{n,k'} > R_{n,k}$ in order to ensure successful decoding of the information by both user k'

and user k. In other words, when $\pi_{n,k'} > \pi_{n,k}$, it is required that the capacity of user k' decoding user k's signal is greater than the capacity of user k decoding his/her own signal [19, 20].

2.2 Problem formulation

The communication scenario considered in this paper is that there are K single-antenna users served by a single-antenna BS, where the LoS link between the BS and the users is not available. To form a wireless communication link between the BS and the K users, an additional array of artificial reflective surfaces is used, consisting of M small reflective elements. The M RISs are sparsely distributed over a surface network of area M_s.

In order to obtain the maximum throughput of the system, we will jointly optimize the topological matrix of the RIS distribution, power allocation, reflection coefficient, channel allocation, and decoding order. The optimized problem can be written as

$$(P1): \max_{\boldsymbol{\delta},\boldsymbol{\pi},\boldsymbol{p},\boldsymbol{\Theta},\boldsymbol{Z}} \sum_{n=1}^{N}\sum_{k=1}^{K} R_{n,k} = \log_2\left(1+\frac{\delta_{n,k}P_{n,k}\left|\boldsymbol{g}_{n,k}^H \boldsymbol{Z}\boldsymbol{\theta}\boldsymbol{f}_n\right|^2}{\left|\boldsymbol{g}_{n,k}^H \boldsymbol{Z}\boldsymbol{\theta}\boldsymbol{f}_n\right|^2 P_{n,k}+\sigma^2}\right) \tag{4}$$

$$s.t. R_{n,k'} \geq R_{n,k}, \left(\pi_{n,k} \leq \pi_{n,k'}\right), \tag{4a}$$

$$R_{n,k'} \geq R_{\min}, \tag{4b}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} \delta_{n,k}P_{n,k} \leq P_{\max}, \tag{4c}$$

$$\left|\Theta_{m,m}\right| \leq 1, \tag{4d}$$

$$\sum_{k=1}^{K} \delta_{n,k} = K_n, \tag{4e}$$

$$\sum_{n=1}^{N} \delta_{n,k} = 1, \tag{4f}$$

$$\pi_n \in \Omega, \tag{4g}$$

$$z_i(z_i - 1) = 0, \tag{4h}$$

$$\|\boldsymbol{Z}\|_1 = M, \tag{4i}$$

where $\boldsymbol{\delta} = \left\{\delta_{1,1}, ..., \delta_{N,K}\right\}$ is the channel assignment indication vector, with $\delta_{n,k} \in \{0,1\}$ indicating whether user k is assigned to channel n. $\boldsymbol{p} = \left\{p_{1,1}, ..., p_{N,K}\right\}$ is the power allocation vector, denotes the power assigned to the nth channel for user k.

$\boldsymbol{\pi} = \{\pi_{1,1}, ..., \pi_{N,K}\}$ is the decoding order vector, representing the order in which user k is decoded on the nth channel. $\boldsymbol{\Theta}$ denotes the reflection coefficient matrix, and $\boldsymbol{Z}$ denotes the selection matrix for the sparse distribution of the RIS. Specifically, the topology matrix $\boldsymbol{Z}$ has M diagonal elements assigned the value of 1, while the rest of the elements are assigned 0.

The optimization problem for the RIS-based multi-user communication system is subject to several constraints to ensure reliable and efficient operation. First, constraint (4*a*) is introduced to guarantee the success of the SIC decoding process, which is a critical component of the system's operation. Second, constraint (4*b*) specifies the minimum capacity requirement $R_{\min}$ for each user to ensure that all users can achieve a certain level of data rate performance. Third, constraint (4*c*) is imposed to control the total transmit power budget, which is a valuable resource that must be used efficiently to minimize interference and maximize capacity. Fourth, constraint (4*d*) is introduced to determine the reflection coefficients of the RIS, which is a key element of the system's infrastructure. Fifth, constraint (4*e*) is used to assign channels to users while ensuring that each channel is allocated to no more than K_n users. This constraint is critical to balancing the load across different channels and avoiding congestion. Sixth, constraint (4*f*) restricts each user to a single channel to avoid interference and ensure fair allocation of resources. Seventh, constraints (4*g*) and (4*h*) limit the decoding order to a combination set Ω of all possible orders, which is necessary to ensure proper SIC decoding and avoid errors. Finally, constraint (4*i*) is introduced to impose sparse deployment of the irregular RIS.

Specifically, for a given topology $\boldsymbol{Z_0}$, ($P1$) can be deduced that

$$(P2): \max_{\boldsymbol{\delta},\boldsymbol{\pi},\boldsymbol{p},\boldsymbol{\Theta}} \sum_{n=1}^{N}\sum_{k=1}^{K} R_{n,k} \tag{5}$$

$$s.t. R_{n,k'} \geq R_{n,k}, \left(\pi_{n,k} \leq \pi_{n,k'}\right), \tag{5a}$$

$$R_{n,k'} \geq R_{\min}, \tag{5b}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} \delta_{n,k} P_{n,k} \leq P_{\max}, \tag{5c}$$

$$\left|\Theta_{m,m}\right| \leq 1, \tag{5d}$$

$$\sum_{k=1}^{K} \delta_{n,k} = K_n, \tag{5e}$$

$$\sum_{n=1}^{N} \delta_{n,k} = 1, \tag{5f}$$

$$\pi_n \in \Omega, \tag{5g}$$

$$\boldsymbol{Z} = \boldsymbol{Z_0}, \tag{5h}$$

3 Joint optimization algorithm proposed for the RIS-Noma system

In order to address the optimization problem ($P1$), we propose a four-step algorithm that aims to achieve an efficient and effective solution. The proposed algorithm consists of the following steps. Firstly, a Simulated Annealing Algorithm is utilized to obtain a sparse deployment of the RIS elements, which can significantly reduce the computational complexity of the subsequent optimization steps. After obtaining the topological matrix Z, it is fixed for the remaining steps. Secondly, an integer optimization algorithm with successive iterations of immobile points is used to solve the power allocation problem. Thirdly, the reflection coefficient matrix is simplified and optimized by constructing appropriate inequalities. Finally, a genetic algorithm is employed to optimize the channel assignment. The proposed algorithm is expected to provide a computationally efficient and effective solution to the optimization problem ($P1$).

3.1 SA-based sparse deployment of RIS

In [2], inspired by [21], they optimized the topology matrix $\boldsymbol{Z}$ using Tabu Search Algorithm. Under the influence of [2], we attempted another widely used metaheuristic algorithm—Simulated Annealing Algorithm [22]. Although this algorithm may not necessarily obtain the optimal solution, it can obtain an approximate optimal solution within a limited time. Taking into account the impact of irregular RIS layouts on system performance in practical scenarios, it is important to note that the relationship is not necessarily monotonic, and multiple local optima may exist. The Simulated Annealing Algorithm offers a potential solution to avoid getting trapped in local optima and has a higher likelihood of approaching the global optimum compared to the Tabu Search Algorithm. Additionally, the Simulated Annealing Algorithm requires fewer algorithm parameters, which makes the overall system implementation and adjustment more manageable.

The specific implementation process of the algorithm is as follows: Firstly, the diagonal elements of the topology matrix $\boldsymbol{Z}$ of the RIS are randomly initialized to obtain $\boldsymbol{Z_0}$, where the number of ones is M and the number of zeros is $M_s - M$. Then, the system performs multiple iterations, and the number of iterations is affected by factors such as the initial temperature, target temperature, cooling schedule, and temperature chain length. In the ith iteration, the RIS topology matrix is $\boldsymbol{Z_i}$, and we randomly exchange the positions of zeros and ones within the diagonal elements of $\boldsymbol{Z_i}$ in a small area. The number and positions of the exchanged elements are randomized. Whether the new RIS topology matrix $\boldsymbol{Z_i}$ is retained is based on the Metropolis criterion and is used for the next iteration. When the number of iterations reaches temperature chain length, the iteration count is reset to zero, and the current temperature is updated and compared with the target temperature until the current temperature is lower than the target temperature, and the iteration ends. At this point, the obtained RIS topology matrix $\boldsymbol{Z}$ is a sub-optimal solution.

Algorithm 1 Irregular RIS Topology Optimization Algorithm

Require: $T_0 > T_{end} > 0$, $0 < \alpha < 1$, $L > 0$
Ensure: $y = \sum_{n=1}^{N} \sum_{k=1}^{K} R_{n,k}$

```
Initialize an initial temperature of T_0, a target temperature of T_end, a cooling schedule of α, and a temperature chain length of L
T = T_0, l = 0, Z = Z_0
if T > T_end then
    if l < L then
        Randomly swap the diagonal elements of matrix Z
        if min(1, e^{-(y-y')/T}) > random(0, 1) then
            Z ⇐ Z'
        end if
        l ⇐ l + 1
    end if
    l ⇐ 0
    T ⇐ T * α
end if
```

3.2 Joint power allocation and reflection coefficient design

Through the solution just now, we have obtained the sub-optimal matrix $\boldsymbol{Z_0}$ of $\boldsymbol{Z}$, and after bringing $\boldsymbol{Z_0}$ into $\boldsymbol{Z}$, we only need to solve the optimization problem of traditional RIS-NOMA.

Let k_n denote the index of the kth user decoded on the nth channel. To optimize the transmit power and the coefficient matrix of the RIS, we first assign channels and specify the decoding order. Then, the capacity $R_{n,kn}$ can be expressed as follows:

$$R_{n,k_n} = \log_2\left(1 + \frac{p_{n,k_n}\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2}{\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 P_{n,k} + \sigma^2}\right) \tag{6}$$

where P_{n,k_n} denotes the total power of all users on the nth channel from the $(k_n + 1)$th user to the K_nth user while communicating at the same time, so $P_{n,k_n} = \sum_{i=k_n+1}^{K_n} p_{n,i_n}$.

Moreover, we can simplify constraint (5a) and rewrite it as

$$\left|\boldsymbol{g}^H{}_{n,k_n'}\boldsymbol{Z}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 - \left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 \geq 0, k_n' > k_n \tag{7}$$

Combining the given data and (6), (P2) can be equivalently transformed into the following problem

$$(P3): \max_{\boldsymbol{p},\boldsymbol{\Theta}} \sum_{n=1}^{N} \sum_{k_n=1}^{K_n} \log_2\left(1 + \frac{p_{n,k_n}\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2}{\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 P_{n,k} + \sigma^2}\right) \tag{8}$$

$$s.t. \log_2\left(1 + \frac{p_{n,k_n}\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2}{\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 P_{n,k} + \sigma^2}\right) \geq R_{\min}, \tag{8a}$$

$$\sum_{n=1}^{N}\sum_{k_n=1}^{K_n} p_{n,k_n} \leq P_{\max}, \tag{8b}$$

$$(5d), (7), \tag{8c}$$

At this point, the variables $\boldsymbol{p}$ and $\boldsymbol{\Theta}$ are still coupled, and ($P3$) is non-convex. Since we cannot optimize these two variables directly, we need to optimize $\boldsymbol{p}$ and $\boldsymbol{\Theta}$ separately and optimization problem ($P3$) can be split into two subproblems

$$(P3.1): \max_{\boldsymbol{p}} \sum_{n=1}^{N}\sum_{k=1}^{K} \log_2\left(1+\frac{p_{n,k_n}\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2}{\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 P_{n,k}+\sigma^2}\right) \tag{9}$$

$$s.t.(8a), (8b), \tag{9a}$$

and

$$(P3.2): \max_{\boldsymbol{\Theta}} \sum_{n=1}^{N}\sum_{k=1}^{K} \log_2\left(1+\frac{p_{n,k_n}\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2}{\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 P_{n,k}+\sigma^2}\right) \tag{10}$$

$$s.t.(5d), (7), \tag{9a}$$

where ($P3.1$) is the optimization problem to find the optimal power allocation vector $\boldsymbol{p}$, amd ($P3.2$) is the optimization problem to find the optimal reflection coefficient matrix $\boldsymbol{\Theta}$.

3.2.1 Derivation of optimization steps for (P3.1)

To solve the subproblem ($P3.1$), we define a new set of variables $X = \{x_{1,k_1}, x_{1,k_n}, \ldots, x_{2,k_1}, x_{2,k_n}, \ldots, x_{n,k_1}, \ldots, x_{n,k_n}\}$, where each x_{n,k_n} satisfies the following inequality:

$$\frac{p_{n,k_n}\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2}{\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 P_{n,k}+\sigma^2} \geq x_{n,k_n} \tag{11}$$

Performing simplification and rewriting techniques on constraint (11) results in

$$p_{n,k_n}\left|\boldsymbol{g}^H{}_{n,k}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 \geq x_{n,k_n}\left(\left|\boldsymbol{g}^H{}_{n,k}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 P_{n,k}+\sigma^2\right) \tag{12}$$

$$p_{n,k_n} \geq x_{n,k_n}P_{n,k} + \left(\frac{\sigma^2}{\left|\boldsymbol{g}^H{}_{n,k}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2}\right) \tag{13}$$

At this juncture, the right-hand side of inequality (15) exhibits quasi-concave; however, the constraint (15) remains non-convex. To address the non-convexity of the problem, a convex upper bound approximation can be utilized [23].

Based on the mean value inequality $ab \geq 2\sqrt{ab}$, we can construct two functions $f(x,y) = \frac{\alpha}{2}x^2 + \frac{1}{2\alpha}y^2(\alpha > 0)$ and $g(x,y) = xy$, respectively. For any x and y, there is always $f(x,y) \geq g(x,y)$, when $\alpha = \frac{y}{x}$, $f(x,y) = g(x,y)$, and the gradients are the same. In other words, $f(x, y)$ is an upper bound on $g(x, y)$.

Based on the above functions, we can construct

$$x_{n,k_n}P_{n,k} \leq \frac{1}{2\alpha}P^2{}_{n,k} + \frac{\alpha}{2}x^2{}_{n,k_n} \tag{14}$$

where α is the fixed-point. Take the equal conditions as $\alpha = \frac{P_{n,k}}{x_{n,k_n}}$. After the ith iteration update, the fixed point $\alpha(t) = \frac{P_{n,k}(t-1)}{x_{n,k_n}(t-1)}$.

With the above analysis, the constraint (13) can be written

$$p_{n,k_n} \geq \frac{1}{2\alpha(t-1)}P^2_{n,k_n} + \frac{\alpha(t-1)}{2}x^2_{n,k_n} + x_{n,k_n}\left(\frac{\sigma^2}{\left|\boldsymbol{g}^H{}_{n,k}\boldsymbol{Z_0}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2}\right) \tag{15}$$

By iterative solution, we can rewrite (P3.1) as

$$(P4): \max_{\boldsymbol{X},\boldsymbol{p}} \sum_{n=1}^{N}\sum_{k=1}^{K}\log_2\left(1 + x_{n,k_n}\right) \tag{16}$$

$$s.t.(8a), (8b), (15), \tag{16a}$$

The optimization problem (P4) is a convex function and is solved using the CVX toolkit [24].

3.2.2 Derivation of optimization steps for (P3.2)

Before optimizing (P3.2), we first simplify $\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2$ as

$$\left|\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z}\boldsymbol{\Theta}\boldsymbol{f}_n\right|^2 = \text{real}(\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z}\boldsymbol{\Theta}\boldsymbol{f}_n)^2 + \text{imag}(\boldsymbol{g}^H{}_{n,k_n}\boldsymbol{Z}\boldsymbol{\Theta}\boldsymbol{f}_n)^2 \tag{17}$$

For ease of notation, we introduce $\boldsymbol{H}_{n,k_n}$ and $\boldsymbol{e_\theta}$ to represent the variables $\boldsymbol{g}^H_{n,k_n}\boldsymbol{Z}\text{diag}\{\boldsymbol{f}_n\}$ and $[\lambda_1 e^{j\theta_1}\ \ldots\ \lambda_{N_s}e^{j\theta_{M_s}}]^T$, respectively. And we further define r_{n,k_n} for $\text{real}(\boldsymbol{H}_{n,k_n}\boldsymbol{e_\theta})$ and i_{n,k_n} for $\text{imag}(\boldsymbol{H}_{n,k_n}\boldsymbol{e_\theta})$.

Here, (P3.2) can be rewritten as

$$(P5): \textit{Find}\ \boldsymbol{e_\theta} \tag{18}$$

$$s.t. r^2_{n,k'_n} + i^2_{n,k'_n} > r^2_{n,k_n} + i^2_{n,k_n}, k'_n > k_n, \tag{18a}$$

$$r^2_{n,k_n} + i^2_{n,k_n} \geq (r^2_{n,k_n} + i^2_{n,k_n})\frac{x_{n,k_n}P_{n,k_n}}{p_{n,k_n}} + x_{n,k_n}\sigma^2, \tag{18b}$$

$$|\boldsymbol{e_\theta}(t)| \leq 1, \tag{18c}$$

$$r_{n,k_n} = \text{real}(\boldsymbol{H}_{n,k_n}\boldsymbol{e_\theta}), \tag{18d}$$

$$i_{n,k_n} = \text{imag}(\boldsymbol{H}_{n,k_n}\boldsymbol{e}_{\boldsymbol{\theta}}), \tag{18e}$$

where $\boldsymbol{e}_{\boldsymbol{\theta}}(t)$ represents the tth iteration of $\boldsymbol{e}_{\boldsymbol{\theta}}$.

The constraints (18a) and (18b) in problem ($P5$) are non-convex, which means that we cannot directly use CVX to optimize the problem. However, we can use the successive convex approximation (SCA) method to rewrite these non-convex constraints into a form that can be handled by CVX.

Based on the sum-of-squares non-negativity theorem, we can formulate an inequality as $(a-a')^2+(b-b')^2 \geq 0$. By bringing r_{n,k_n} and i_{n,k_n} into a and b, respectively, the inequality can be written as $(r_{n,k_n}-r'_{n,k_n})^2+(i_{n,k_n}-i'_{n,k_n})^2 \geq 0$. The point $(r'n,k_n,i'n,k_n)$ in equation represents a first-order approximation of the point (r_{n,k_n},i_{n,k_n}), which can be updated iteratively.

Simplify this inequality as

$$r^2_{n,k_n}+i^2_{n,k_n} \geq -{r'_{n,k_n}}^2-{i'_{n,k_n}}^2+2r_{n,k_n}r'_{n,k_n}+2i_{n,k_n}i'_{n,k_n} \tag{19}$$

$$r^2_{n,k_n}+i^2_{n,k_n} \geq {r'_{n,k_n}}^2+{i'_{n,k_n}}^2+2r'_{n,k_n}(r_{n,k_n}-r'_{n,k_n})+2i'_{n,k_n}(i_{n,k_n}-i'_{n,k_n}) \tag{20}$$

During the tth iteration, the point $(r'n,k_n,i'n,k_n)$ can be updated as

$$r'_{n,k_n}(t) = \text{real}(\boldsymbol{H}_{n,k_n}\boldsymbol{e}_{\boldsymbol{\theta}}(t-1)), \tag{21}$$

$$i'_{n,k_n}(t) = \text{imag}(\boldsymbol{H}_{n,k_n}\boldsymbol{e}_{\boldsymbol{\theta}}(t-1)), \tag{22}$$

Assume that the function $F(r_{n,k_n},i_{n,k_n})={r'_{n,k_n}}^2+{i'_{n,k_n}}^2+2r'_{n,k_n}(r_{n,k_n}-r'_{n,k_n})+2i'_{n,k_n}(i_{n,k_n}-i'_{n,k_n})$, we can derive the inequality $r^2_{n,k_n}+i^2_{n,k_n} \geq F(r_{n,k_n},i_{n,k_n})$.

The constraints (18a) and (18b) can be simplified as

$$F(r_{n,k'_n},i_{n,k'_n}) > r^2_{n,k_n}+i^2_{n,k_n}, k'_n > k_n, \tag{23}$$

$$F(r_{n,k_n},i_{n,k_n}) \geq (r^2_{n,k_n}+i^2_{n,k_n})\frac{x_{n,k_n}P_{n,k_n}}{p_{n,k_n}}+x_{n,k_n}\sigma^2, \tag{24}$$

After iteration the optimization problem ($P5$) can be written as

$$(P6): \text{Find } \boldsymbol{e}_{\boldsymbol{\theta}} \tag{25}$$

$$s.t.(18c),(18d),(18e),(23),(24), \tag{25e}$$

The optimization problem ($P6$) is a convex function, which can be solved using the CVX toolkit.

3.3 Optimize the allocation of channels

In this subsection, we will discuss the optimization of the channel assignment problem in an irregular RIS-assisted NOMA system. Given the topology matrix of the irregular RIS, power allocation, reflection coefficient, and decoding order, ($P1$) can be rewritten as

$$(P7): \max_{\delta} \sum_{n=1}^{N} \sum_{k=1}^{K} R_{n,k} \tag{26}$$

$$s.t.(5a), (5b), (5e), (5f), \tag{26a}$$

To optimize problem (*P*7), we used a genetic algorithm.

First, we introduce the set $\mathcal{H}$, which represents the set of all channels and can be written $\mathcal{H} = \{\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_N\}$, where $\mathcal{H}_n$ denotes the set of users on channel n. And $\mathcal{H}_1 \cup \mathcal{H}_2 \cup ... \cup \mathcal{H}_N = \mathcal{K}$, where $\mathcal{K}$ denotes the set of all users. Combining constraints (5*a*) and (5*b*), we define

$$|\mathcal{H}| = N \tag{27}$$

$$|\mathcal{H}_n| = K_n \tag{28}$$

$$\forall k \in \mathcal{K}, \exists! \mathcal{H}_n \text{ such that } k \in \mathcal{H}_n \tag{29}$$

where constraint (27) ensures the existence of N channels, constraint (28) ensures that channel n is allocated to K_n users, and constraint (29) ensures that each user k is allocated to one and only one channel.

With the channel allocation representation established, all feasible channel allocation schemes can be represented by the set $\mathcal{H}$. For any $\mathcal{H}$, it corresponds to a valid channel allocation scheme. Additionally, there exists a one-to-one correspondence between $\mathcal{H}$ and the channel allocation scheme, satisfying the normality of the coding strategy.

To optimize the NOMA channel allocation problem, we first randomly generate an initial set of feasible solutions $\mathcal{H}(0)$ that satisfy the three constraints, including 10 randomly generated feasible solutions $\mathcal{H}$.

After generating the initial set of solutions, we enter the iterative process. In each iteration, we select the solution in the current set with the maximum system throughput and retain it while discarding the others.

Then, we randomly modify the matching relationships between some users and channels in the retained solution set $\mathcal{H}$, resulting in 10 new feasible solutions that satisfy the constraints.

Additionally, we generate a completely random solution that satisfies only the constraints, and is independent of $\mathcal{H}$, where increases the diversity of the solutions and helps to avoid getting trapped in local optima. By adding a randomly generated solution $\mathcal{H}$ to $\mathcal{H}(t)$, we can further increase the diversity of the solution set $\mathcal{H}(t)$ and help avoid premature convergence.

After reaching the specified number of iterations T, we select the solution $\mathcal{H}$ in the set $\mathcal{H}(T)$ with the maximum system throughput as the optimal solution.

Algorithm 2 Allocation Of Channels Optimization Algorithm

Ensure: $y = \sum_{n=1}^{N} \sum_{k=1}^{K} R_{n,k}$

1: **Initialize** an initial set of feasible solutions $\mathcal{H}(0)$, number of iterations T
2: $t = 0$
3: **if** $t < T$ **then**
4: $\mathcal{H}^* = \arg\max_{\mathcal{H}_n \in \mathcal{H}(t)} y$
5: Randomly swap the elements of the matrix $\mathcal{H}^*$ generates 10 candidate solutions, each of which is included in the solution set $\mathcal{H}(t+1)$. In addition, a completely random and independent set of $\mathcal{H}^*$ is generated and added to the solution set $\mathcal{H}(t+1)$.
6: $t \Leftarrow t+1$
7: **end if**

4 Simulation results

In this subsection, we will build a MATLAB simulation platform to test the throughput maximization algorithm of "Irregular Reconfigurable Intelligent Surface Assisted NOMA Systems" designed in this paper. In this subsection, two aspects of the simulation are tested: The first aspect is the comparison of the algorithm with the capacity of conventional RIS and non-regular RIS, NOMA system and OMA system from two aspects: different SNRs and different number of reflecting elements. The second aspect is to compare the performance of four algorithms, namely, the traversal algorithm, the Simulated Annealing Algorithm, the tabu search algorithm, and the random assignment algorithm, from the perspective of their convergence performance.

The first is the simulation of the system capacity performance, and a brief description of the simulation scheme is given. In this simulation, it is assumed that there is a BS with a total of K single-antenna users, where the LoS link between the BS and the users is not available, and the communication between the BS and the users is formed by an additional array of artificial reflective surfaces (consisting of M small reflective surfaces) to form a wireless communication link, and the small-scale fading of the link between the BS and the RIS and the link between the RIS and the users obeys the Rayleigh distribution, the channel responses are independent of each other among users and the large-scale fading obeys the free-space loss with a loss factor of 2. Other channel configuration parameters are shown in Table 1.

Table 1 Channel capacity simulation parameters configuration

Parameter name	Parameter value
K (Number of users)	4
M (Number of RIS)	32
Noise power (dBm)	− 80
RIS to base station distance	50
RIS to user distance	2
Number of Monte Carlo iterations	100
Base station transmit power (dBm)	$-10, -5, 0, 5, \ldots, 30$

In Fig. 2, the system capacity of Traditional regular RIS-Assisted NOMA System, Proposed irregular RIS-Assisted NOMA System, Traditional regular RIS-Assisted OMA System and Proposed irregular RIS-Assisted OMA System. The system capacity of these four systems varies with the transmitting power of the BS. It can be seen from Fig. 2 that the system capacity of these four systems increases linearly with the increase in the transmitting power of the BS, and Proposed irregular RIS-Assisted NOMA System has the best capacity performance, followed by Traditional regular RIS-Assisted NOMA System. Traditional regular RIS-Assisted NOMA System, whose performance is about 2.5 dB worse, and then Proposed irregular RIS-Assisted OMA System, whose performance is about 6 dB worse. Based on the above results, it can be seen that irregular RIS has a higher performance than traditional regular RIS in both NOMA and OMA systems. Irregular RIS is about 3 dB better than the traditional regular RIS, which indicates that the limited reflecting surface can be used more efficiently by designing the location of the reflecting surface.

Compared with the traditional regular RIS-assisted OMA system, the irregular RIS-assisted NOMA system proposed in this paper has achieved a performance improvement of about 30%, and even the weak irregular RIS-assisted OMA system has a maximum throughput improvement of about 20%. This highlights the great superiority of irregular RIS.

In Fig. 3, the effect of the total reflecting array size on the system capacity under the irregular RIS is analyzed for a total number of reflecting surface elements of 16. The capacity performance of the four systems is not different from that in Fig. 2, but the performance of the irregular RIS system increases as the size of the reflecting array increases, indicating that increasing the size of the reflecting array can effectively increase the capacity performance of the system.

However, RIS elements with values of 70, 90, and 100 show slight decreases to varying degrees. This is because when the total size of the reflection array reaches four

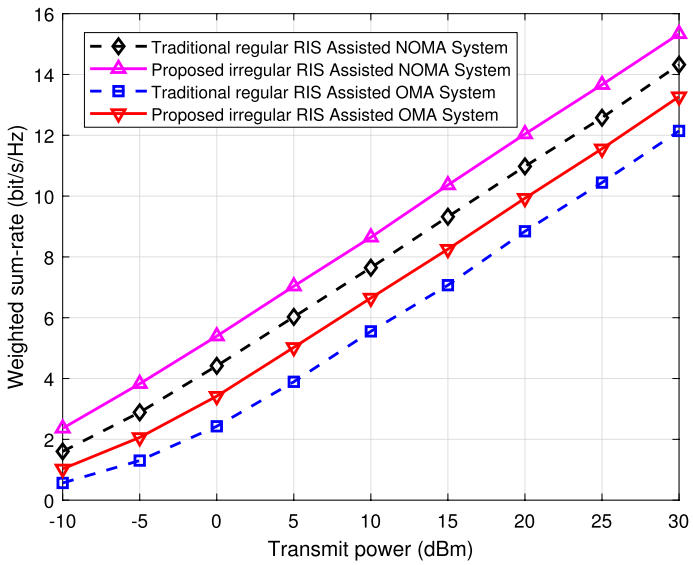

Fig. 2 Maximum capacity in different situations

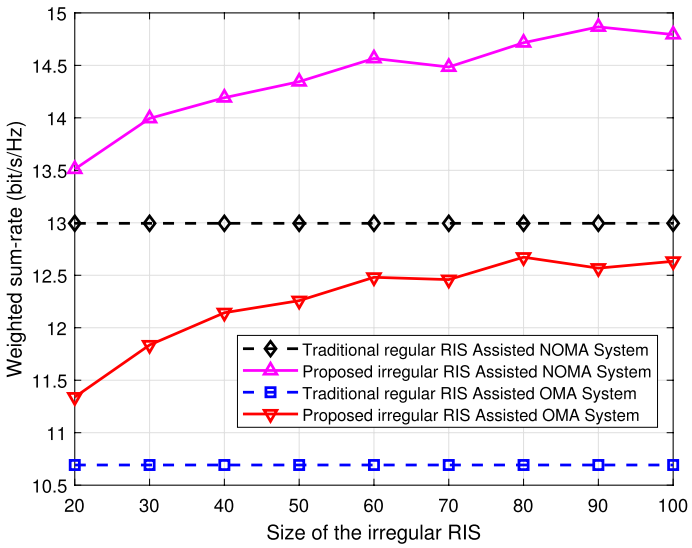

Fig. 3 Maximum capacity in different situations

times the effective area of the reflective elements, the capacity performance of the system reaches a convergence point, suggesting that there is a certain upper limit to the performance improvement achieved by increasing the size of the reflection array, and further increasing the size of the reflection array becomes less significant.

This phenomenon becomes even more pronounced in practical applications due to an increase in the number of irregular RIS elements, which leads to multiple signal reflections and introduces additional path loss and interference. The increase in channel fading and interference results in a decrease in both the maximum system performance and rate. Moreover, in NOMA systems, larger RIS configurations may give rise to the issue of coherent signal superposition. When multiple reflected signals from the RIS interfere coherently with the primary signal, the interference level rises, thereby reducing the maximum system performance and rate.

This highlights the necessity of irregular RIS, when the RIS size is continuously increased, the performance improvement is no longer obvious, at this time replacing the traditional regular RIS with irregular RIS can greatly improve the maximum throughput of the system. From the perspective of installation cost and RIS cost, the construction of irregular RIS is reduced in size compared with traditional regular RIS under the same performance, and the lower cost undoubtedly reflects another advantage of irregular RIS.

Combining the results of Figs. 2 and 3, it is easy to conclude that the use of irregular RIS can effectively improve the overall system capacity performance without increasing the number of reflecting surface elements and the corresponding phase control unit, and the larger the reflecting array size, the more obvious the improvement is, and the overall performance improvement is about 3–5 dB. The overall performance improvement is about 3–5 dB, which fully proves the superiority of irregular RIS.

Apart from comparing the system's maximum capacity and speed, we should also take into account the complexity and feasibility of the system. Next is an analysis of

the performance of four different search algorithms. The traversal search algorithm guarantees finding the global optimal solution and represents the upper limit of performance. However, it has high computational complexity and is suitable for small-scale problems. The Simulated Annealing Algorithm possesses global search ability, making it suitable for non-convex optimization problems, but its performance is sensitive to parameter selection. The Tabu Search Algorithm excels in local search and convergence to local optima but may fall into them. The random assignment algorithm, representing the lower limit of performance, is simple but offers poor results as it lacks problem-specific considerations. For a fixed number of RIS elements and the number of elements in the entire panel, focusing on the performance of the Simulated Annealing Algorithm and the Tabu Search Algorithm as a function of iteration and percentage compared to the optimal case. The simulation scheme is basically the same as in Figs. 2 and 3, The four different optimization search algorithms, namely, Traverse Algorithm, Simulated annealing Algorithm, Tabu Search Algorithm and Random Number Algorithm, are applied to the search of the reflecting surface elements of the irregular RIS mentioned in Figs. 2 and 3. The optimization performance and complexity performance simulation scenarios of the four algorithms in searching for irregular intelligent reflecting surface elements are basically the same as the one in which a BS is assumed to have a total of K single-antenna users, where the LoS link between the BS and the users is not available, and the communication between the BS and the users is carried out through an additional artificially erected reflecting surface element array (composed of N small reflecting surfaces) constitutes the wireless communication link, and the small-scale fading obeys the Rayleigh distribution in the link between the BS to the RIS and the link between the RIS to the user, and the large-scale fading obeys the free-space loss, and the loss factor is 2. Other channel configuration parameters are shown in Table 1.

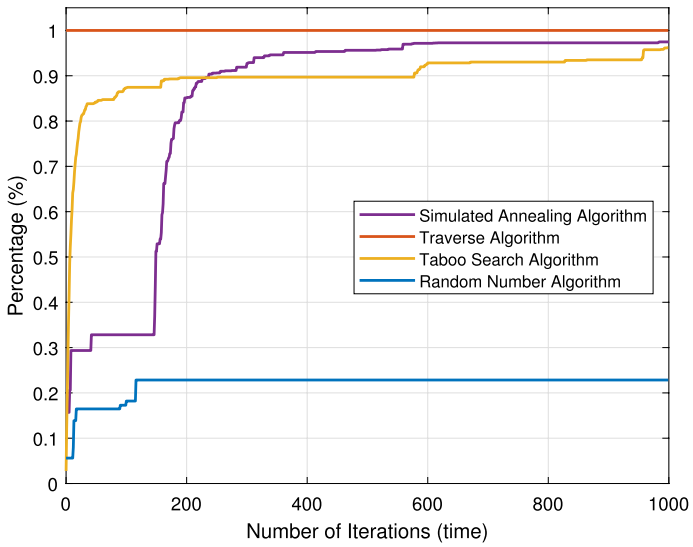

Fig. 4 Comparison of sparse distribution feasibility

Figure 4 illustrates the relationship between the number of iterations of the four algorithms and the percentage of time spent by the optimal algorithm. Although time spent cannot fully replace time complexity, it can provide some insight into the practical performance of the four algorithms. As shown in Fig. 4, the Simulated Annealing Algorithm requires significantly less time than the Optimal Algorithm and the Tabu Search Algorithm in the first 500 iterations, taking only 20% of the time of the optimal algorithm. The forbidden search algorithm already takes 60% of the time of the optimal algorithm at this point. However, after 500 iterations, the Simulated Annealing Algorithm consumes significantly more time per iteration, and at 912 iterations, the total time spent exceeds that of the Tabu Search Algorithm and approaches that of the Optimal Algorithm. These results indicate that the Simulated Annealing Algorithm requires very little time for a small number of iterations, but as the number of iterations increases, it gradually approaches and surpasses the Tabu Search Algorithm. It should be noted that since the time to generate random numbers is negligible compared to the time spent on iterations, the time spent by the Random Number Algorithm is negligible compared to the time spent by the Optimal Algorithm. This is evident in Fig. 4, where the time spent by the Random Number Algorithm is consistently 0. It is important to note that the number of iterations in both the Simulated Annealing Algorithm and the Tabu Search Algorithm is user-defined. In this simulation, a higher number of iterations was set to simulate complex scenarios. This resulted in the Simulated Annealing Algorithm requiring a substantial amount of time, approaching the execution time of the Optimal Algorithm, and even showing a tendency to exceed it. However, this does not imply that the complexity of the Simulated Annealing Algorithm is higher than that of the Optimal Algorithm.

Figure 5 presents the percentage relationship between the number of iterations of the four algorithms and the performance of the Optimal Algorithm. As shown in the figure, when the number of iterations reaches 100, the Tabu Search Algorithm has achieved 88% of the performance of the Optimal Algorithm, and by 200 iterations, it has reached

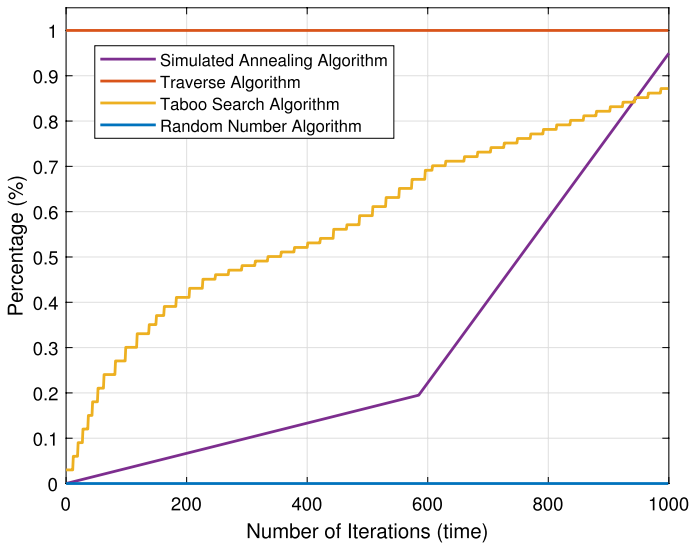

Fig. 5 Comparison of sparse distribution complexity

90%. The Simulated Annealing Algorithm also achieves 90% of the Optimal Algorithm's performance at 220 iterations and surpasses the Tabu Search Algorithm at the 226th iteration. By 300 iterations, the Simulated Annealing Algorithm has reached 98% performance, and the performance of both algorithms is very close to that of the Optimal Algorithm. In contrast, the performance of the Stochastic Algorithm is only 20%-30% of the Optimal Algorithm's level, making it of little practical value despite its low complexity.

The results in Fig. 4 demonstrate that at 300 iterations, the Simulated Annealing Algorithm requires only 10% of the operation time of the Optimal Algorithm, while achieving a performance close to 98% of the Optimal Algorithm. Combining the results shown in Figs. 4 and 5, we can conclude that the Simulated Annealing Algorithm generally outperforms the Tabu Search Algorithm, although its performance may be inferior to the Tabu Search Algorithm in a few iterations (less than 226), and its operation time may exceed that of the Tabu Search Algorithm after many iterations (more than 912). Nonetheless, the Simulated Annealing Algorithm can effectively reduce the overall complexity of the system while largely maintaining its overall performance.

Despite only providing a sub-optimal solution for the current scenario, the Simulated Annealing Algorithm can achieve good results after a short iteration time. Compared to traditional traversal algorithms, the Simulated Annealing Algorithm achieves lower complexity while still producing satisfactory results. Additionally, the results are flexible and adjustable. In situations where the environment is variable, and the system cannot iterate many times, the algorithm can quickly achieve over 90% system performance with a short iteration time, which is a significant improvement compared to the ordinary RIS-assisted OMA system. Conversely, if the environment is stable and sufficient time is available, the algorithm can achieve a near-optimal solution by increasing the iteration time. The flexibility of the number of iterations, coupled with its improved performance, renders the model more suitable for practical scenarios and expands its application prospects.

Based on the results in Figs. 2, 3, 4, and 5 above, it can be seen that Irregular Intelligent Reflecting Surface Assisted NOMA Systems and Irregular Intelligent Reflecting Surface Assisted OMA Systems have a better performance than Traditional Intelligent Reflecting Surface Assisted NOMA Systems and Traditional Intelligent Reflecting Surface Assisted OMA Systems. Systems have a significant performance improvement over Traditional Intelligent Reflecting Surface Assisted NOMA Systems and Traditional Intelligent Reflecting Surface Assisted OMA Systems. The combination of the Tabu Search Algorithm mentioned in this paper can reduce the overall complexity of the system by more than 50% without affecting the overall performance, which fully guarantees the possibility of practical application of Irregular Intelligent Reflecting Surface.

5 Conclusion

In this paper, the optimization of an RIS-assisted NOMA system to maximize system throughput was investigated. The proposed approach co-optimized the beamforming of the BS, passive phase shifting of the RIS, and channel assignment to achieve this goal. A new step-by-step optimization algorithm was introduced to solve the coupling optimization problem. Through simulation, the proposed algorithm was shown to be effective

and capable of real-time adjustment of channel allocation and RIS element distribution to achieve higher system throughput with lower computational effort.

Furthermore, the optimization scheme of RIS element distribution mentioned in the paper achieved sub-optimal results while reducing the overall complexity of the system by more than 50% compared to other methods and scenarios. The combination of the Tabu Search Algorithm mentioned in the paper further reduced the computational complexity of the system. Overall, the paper provided a promising approach to maximizing the throughput of an RIS-assisted NOMA system with practical benefits such as reduced complexity and improved efficiency. The proposed step-by-step optimization algorithm effectively optimized the beamforming of the BS, passive phase shifting of the RIS, and channel assignment in a co-optimized manner, leading to improved system throughput. Additionally, the proposed optimization scheme for RIS element distribution significantly reduced the computational complexity of the system. The combination of these two approaches resulted in a practical and efficient solution for optimizing RIS-assisted NOMA systems. The results of the simulations demonstrated the effectiveness of the proposed approach and suggested that it had the potential to significantly improve the performance of current wireless communication systems.

The future research directions encompass the inclusion of LoS links in simulation scenarios to further enhance system performance. This integration will facilitate a more accurate emulation of real-world communication environments, providing a realistic foundation for the design and optimization of IRS systems. Furthermore, it is imperative to explore methods for dynamically adjusting the number of IRS elements and optimizing the irregular IRS topology matrix with an iterative upper limit tailored to specific scenarios. The flexible adjustment of IRS quantities and optimization of the topology matrix will enable precise system configuration and resource utilization, thereby elevating system performance and efficiency.

In forthcoming investigations, significant efforts should also be dedicated to minimizing the number of iterations to enhance system efficiency. To this end, novel algorithms and optimization strategies can be explored to reduce computational complexity and improve convergence speed. By reducing the number of iterations, the process of system design and optimization can be accelerated while concurrently reducing the requirements for computational resources and time, consequently advancing the practicability and feasibility of irregular IRS-assisted NOMA systems in practical scenarios.

Similar to the network security concerns highlighted in [25, 26], our future research endeavors will also focus on investigating the network security aspects of irregular IRS-assisted NOMA systems.

Author contributions

WZ made significant contributions throughout the entire research process as the lead author, driving the progress of the paper. He played a key role in designing and implementing the study, demonstrating his dedication and expertise in the field. LW, HM, and ZW provided valuable assistance in the code emulation work, making important contributions to the project. They displayed technical proficiency and commitment, which greatly enhanced the overall quality of the research. WW played a crucial role in guiding the research directions and establishing the conceptual framework. His deep domain knowledge and insightful perspectives were instrumental in shaping the study, ensuring its relevance and significance within the broader research landscape.

Funding

This work was partially funded by the National Key Research and Development Program of China under Grant 2020YFB1807602 and the National Natural Science Foundation of China under Grant 6227126.

Availability of data and materials
Not applicable.

Code availability
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no Competing interests.

Received: 21 June 2023 Accepted: 23 October 2023
Published online: 01 November 2023

References

1. Z. Ding, H.V. Poo, A simple design of IRS-NOMA transmission. IEEE Commun. Lett. **24**(5), 1119–1123 (2020). https://doi.org/10.1109/LCOMM.2020.2974196
2. M. Shehab, B.S. Ciftler, T. Khattab, M.M. Abdallah, D. Trinchero, Deep reinforcement learning powered IRS-assisted downlink NOMA. IEEE Open J. Commun. Soc. **3**, 729–739 (2022). https://doi.org/10.1109/OJCOMS.2022.3165590
3. E.-K. Hong, I. Lee, B. Shim, Y.-C. Ko, S.-H. Kim, S. Pack, K. Lee, S. Kim, J.-H. Kim, Y. Shin, Y. Kim, H. Jung, 6g r &d vision: requirements and candidate technologies. J. Commun. Netw. **24**(2), 232–245 (2022). https://doi.org/10.23919/JCN.2022.000015
4. S. Yrjölä, P. Ahokangas, M. Matinmikko-Blue, Value creation and capture from technology innovation in the 6g era. IEEE Access **10**, 16299–16319 (2022). https://doi.org/10.1109/ACCESS.2022.3149590
5. R. Su, L. Dai, J. Tan, M. Hao, R. MacKenzie, Capacity enhancement for irregular reconfigurable intelligent surface-aided wireless communications, in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, (2020), pp. 1–6. IEEE. https://doi.org/10.1109/GLOBECOM42002.2020.9348273
6. D. Zhang, Q. Wu, M. Cui, G. Zhang, D. Niyato, Throughput maximization for IRS-assisted wireless powered hybrid NOMA and TDMA. IEEE Wirel. Commun. Lett. **10**(9), 1944–1948 (2021). https://doi.org/10.1109/LWC.2021.3087495
7. L. Chuan, C. Qing, L. Xianxu, Uplink NOMA signal transmission with convolutional neural networks approach. J. Syst. Eng. Electr. **31**(5), 890–898 (2020). https://doi.org/10.23919/JSEE.2020.000068
8. W. Zhou, J. Xia, C. Li, L. Fan, A. Nallanathan, Joint precoder, reflection coefficients, and equalizer design for IRS-assisted MIMO systems. IEEE Trans. Commun. **70**(6), 4146–4161 (2022). https://doi.org/10.1109/TCOMM.2022.3168284
9. Z. Ding, R. Schober, H.V. Poor, On the impact of phase shifting designs on IRS-NOMA. IEEE Wirel. Commun. Lett. **9**(10), 1596–1600 (2020). https://doi.org/10.1109/LWC.2020.2991116
10. E. Bjørnson et al., intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying? IEEE Wirel. Commun. Lett. **9**(2), 244–248 (2020). https://doi.org/10.1109/LWC.2019.2950624
11. X. Xie, F. Fang, Z. Ding, Joint optimization of beamforming, phase-shifting and power allocation in a multi-cluster IRS-NOMA network. IEEE Trans. Vehic. Technol. **70**(8), 7705–7717 (2021). https://doi.org/10.1109/TVT.2021.3090255
12. Z. Li, W. Chen, Q. Wu, K. Wang, J. Li, Joint beamforming design and power splitting optimization in IRS-assisted swipt NOMA networks. IEEE Trans. Wirel. Commun. **21**(3), 2019–2033 (2022). https://doi.org/10.1109/TWC.2021.3108901
13. S. Tang, L. Chen, K. He, J. Xia, L. Fan, A. Nallanathan, Computational intelligence and deep learning for next-generation edge-enabled industrial IoT. IEEE Trans. Netw. Sci. Eng. (2022). https://doi.org/10.1109/TNSE.2022.3180632
14. S. Zheng, C. Shen, X. Chen, Design and analysis of uplink and downlink communications for federated learning. IEEE J. Select. Areas Commun. **39**(7), 2150–2167 (2021). https://doi.org/10.1109/JSAC.2020.3041388
15. L. Chen, L. Fan, X. Lei, T.Q. Duong, A. Nallanathan, G.K. Karagiannidis, Relay-assisted federated edge learning: performance analysis and system optimization. IEEE Trans. Commun. **71**(6), 3387–3401 (2023). https://doi.org/10.1109/TCOMM.2023.3263566
16. X. Wang, M. Amin, X. Cao, Analysis and design of optimum sparse array configurations for adaptive beamforming. IEEE Trans. Signal Process. **66**(2), 340–351 (2018). https://doi.org/10.1109/TSP.2017.2760279
17. Y. Yang, B. Zheng, S. Zhang, R. Zhang, Intelligent reflecting surface meets OFDM: protocol design and rate maximization. IEEE Trans. Commun. **68**(7), 4522–4535 (2020)
18. O. Ozdogan, E. Bjornson, E.G. Larsson, Intelligent reflecting surfaces: physics, propagation, and pathloss modeling. IEEE Wirel. Commun. Lett. **9**(5), 581–585 (2020). https://doi.org/10.1109/LWC.2019.2960779
19. Y. Liu, M. Elkashlan, Z. Ding, G.K. Karagiannidis, Fairness of user clustering in MIMO non-orthogonal multiple access systems. IEEE Commun. Lett. **20**(7), 1465–1468 (2016). https://doi.org/10.1109/LCOMM.2016.2559459
20. J. Cui, Y. Liu, Z. Ding, P. Fan, A. Nallanathan, Optimal user scheduling and power allocation for millimeter wave NOMA systems. IEEE Trans. Wirel. Commun. **17**(3), 1502–1517 (2018). https://doi.org/10.1109/TWC.2017.2779504
21. X. Gao, L. Dai, C. Yuen, Z. Wang, Turbo-like beamforming based on tabu search algorithm for millimeter-wave massive MIMO systems. IEEE Trans. Vehic. Technol. **65**(7), 5731–5737 (2016). https://doi.org/10.1109/TVT.2015.2461440

22. B. Aylaj, M. Belkasmi, H. Zouaki, A. Berkani, Degeneration simulated annealing algorithm for combinatorial optimization problems, in *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)* (2015), pp. 570–575. IEEE. https://doi.org/10.1109/ISDA.2015.7489177
23. L.-N. Tran, M.F. Hanif, A. Tolli, M. Juntti, Fast converging algorithm for weighted sum rate maximization in multicell miso downlink. IEEE Signal Process. Lett. **19**(12), 872–875 (2012). https://doi.org/10.1109/LSP.2012.2223211
24. M. Grant, Cvx: Matlab software for disciplined convex programming (2008). http://cvxr.com/cvx
25. Y. Guo, R. Zhao, S. Lai, L. Fan, X. Lei, G.K. Karagiannidis, Distributed machine learning for multiuser mobile edge computing systems. IEEE J. Select. Topics Signal Process. **16**(3), 460–473 (2022). https://doi.org/10.1109/JSTSP.2022.3140660
26. J. Xia, L. Fan, W. Xu, X. Lei, X. Chen, G.K. Karagiannidis, A. Nallanathan, Secure cache-aided multi-relay networks in the presence of multiple eavesdroppers. IEEE Trans. Commun. **67**(11), 7672–7685 (2019). https://doi.org/10.1109/TCOMM.2019.2935047

Publisher's Note