

Optimization of Energy-Constrained IRS-NOMA Using a Complex Circle Manifold Approach

Mahmoud AlaaEldin, *Student Member, IEEE*, Emad Alsusa, *Senior Member, IEEE*, Karim Seddik, *Senior Member, IEEE*, Mohammad Al-Jarrah, *Member, IEEE*, and Constantinos Papadias, *Fellow, IEEE*

Abstract—This work investigates the performance of intelligent reflective surfaces (IRSs) assisted uplink non-orthogonal multiple access (NOMA) in energy-constrained networks. Specifically, we formulate and solve two optimization problems; the first aims at minimizing the sum of users’ transmit power, while the second targets maximizing the system-level energy efficiency (EE). The two problems are solved by jointly optimizing the users’ transmit powers and the beamforming coefficients at the IRS, subject to the users’ individual uplink rate and transmit power constraints. A novel and low-complexity algorithm is developed to optimize the IRS beamforming coefficients by optimizing the objective function over the *complex circle manifold* (CCM). To efficiently optimize the IRS phase shifts over the manifold, the optimization problem is reformulated into a feasibility expansion problem which is reduced to a max-min signal-plus-interference-ratio (SINR). Then, with the aid of a smoothing technique, the exact penalty method is applied to transform the problem from constrained to unconstrained. The proposed solution is compared against three semi-definite programming (SDP)-based benchmarks which are semi-definite relaxation (SDR), SDP-difference of convex (SDP-DC) and sequential rank-one constraint relaxation (SROCR). The results show that the manifold algorithm provides better performance than the SDP-based benchmarks, and at a much lower computational complexity for both the transmit power minimization and EE maximization problems. The results also reveal that IRS-NOMA is only superior to orthogonal multiple access (OMA) when the users’ target achievable rate requirements are relatively high.

Index Terms—Non-orthogonal multiple access, intelligent reflective surfaces, energy-efficient networks, manifold optimization, complex circle manifold, semi-definite relaxation.

I. INTRODUCTION

Recently, intelligent reflective surfaces (IRSs) have emerged as a new technology due to their potential for improving the spectral and energy efficiency of wireless networks [1]–[5]. Generally speaking, an IRS is a planar surface that consists of a large number of adjustable and low-cost passive reflecting elements which reflect incident signals after adjusting their amplitudes and phase shifts. The merit of IRSs lies in their ability to *engineer* the wireless channel to realize different design needs such as signal strengthening and interference mitigation.

As such, integrating IRSs in different wireless applications has been extensively studied in the literature. For example, in

Mahmoud AlaaEldin, Emad Alsusa, and Mohammad Al-Jarrah are with the Electrical and Electronic Engineering Department, University of Manchester, Manchester, UK M13 9PL (e-mail: mahmoud.alaeldin@manchester.ac.uk; e.alsusa@manchester.ac.uk; mohammad.al-jarrah@manchester.ac.uk).

Karim G. Seddik is with the Department of Electronics and Communications Engineering, American University in Cairo, Cairo, Egypt 11835 (e-mail: kseddik@aucegypt.edu).

Constantinos B. Papadias is with the Research, Technology and Innovation Network (RTIN), Alba, The American College of Greece, 153 42 Athens, Greece, and also with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark (e-mail: cpapadias@acg.edu).

[6], the authors studied the application of IRSs in a single cell scenario with a multi-antenna base station (BS) and single-antenna users. They developed alternation minimization algorithms to minimize the total transmit power at the BS by applying a semi-definite relaxation (SDR) approach to optimize the beamforming at the IRS. In [7], the authors proposed efficient algorithms to maximize the energy efficiency (EE) of IRS-assisted downlink communication systems with a multi-antenna BS serving multi-antenna users. More specifically, they developed efficient algorithms based on alternation maximization, gradient descent, and sequential fractional programming to jointly optimize the transmit power allocation at the BS and the phase shifts at the IRS. Furthermore, in [8], a joint power and user association scheme was proposed for a multi-IRS-assisted multi-BS downlink millimeter-Wave system serving multiple users. Other applications of IRSs include unmanned aerial vehicles (UAV) communications [9], physical layer security for covert communications [10], physical layer network coding [11], and wireless mesh back-hauling [12].

Using IRSs in energy-constrained Internet of Things (IoT) systems presents an attractive solution to address the challenge of low uplink data rates, which stem from their limited battery capacity. Since massive connectivity is another crucial requirement for IoT systems, multiple access (MA) techniques play a pivotal role in accommodating a large number of IoT devices. Existing MA techniques can be broadly categorized into two groups: orthogonal multiple access (OMA) and non-orthogonal multiple access (NOMA). The NOMA MA technique stands out as a practical choice for IoT networks, offering the ability to facilitate the access of numerous devices by allowing multiple users to simultaneously use the same spectrum [13], [14]. The merits of NOMA have made it a key technique to enhance spectral efficiency (SE) and EE of wireless communication networks [15]–[17]. Unlike OMA, NOMA can serve multiple users using the same frequency/time/code resources, and applies successive interference cancellation (SIC) at the receiver to reduce co-user interference [17], [18]. In comparison to OMA, integrating NOMA into conventional IoT applications without IRSs has demonstrated several advantages, particularly in applications like Mobile Edge Computing (MEC) [19], and data collection systems [20], [21].

The celebrated merits of NOMA and IRSs make them potential candidates for future wireless networks that can enable various IoT applications. Motivated by this, we investigate the potential of integrating IRSs and NOMA in uplink communications. Specifically, an optimization-based study on the applicability and effectiveness of the introduced system is presented to jointly optimize the users’ transmit powers and IRS phase shifts. The aim is to minimize the power

consumption of battery-based nodes and to maximize the EE of the system while satisfying the quality-of-service (QoS) and transmit power constraints for each node.

A. Related Literature on integrated IRS-NOMA

Much effort has been devoted to integrating IRSs with NOMA. Prior research on this topic can be divided into two categories: downlink IRS-NOMA and uplink IRS-NOMA. While most of the work on this topic is related to downlink IRS-NOMA, there exist only a few articles that consider the uplink scenario. Therefore, in this paper, we are interested in the optimization of the latter scenario by jointly optimizing the users' transmit powers and IRS phase shifts. We first focus on reviewing optimized uplink IRS-NOMA systems, then, for the sake of completeness, we review relevant works on the downlink scenario.

Early work on uplink IRS-NOMA optimization can be found in [22], which considered the users' sum rate maximization under individual QoS and power constraints, assuming perfect knowledge of the IRS-BS and IRS-users channels¹. To achieve their objectives, the authors converted the optimization problem to a semi-definite programming (SDP) problem, and then used SDR to optimize the IRS reflecting coefficients. Besides, the authors in [28] studied the max-min fairness secrecy rate of a two-user IRS assisted uplink NOMA scenario. The solution of the formulated optimization problem includes the conversion of the passive beamforming problem at IRS to SDP followed by employing sequential rank-one constraint relaxation (SROCR) to get a rank-one solution for SDP. In addition, a two-step optimization scheme is devised in [29] for solving the sum rate maximization problem of an active-IRS assisted uplink NOMA scenario. A fixed point iteration (FPI) method has been employed to optimize the phase shifts, which has performance degradation as the number of IRS reflectors increases. In [30], the max-min fairness rate optimization for an IRS assisted uplink NOMA system with receive beamforming at BS was studied. Successive convex approximation (SCA) was used to solve the passive beamforming problem at IRS using second order cone programming (SOCP) in order to avoid the high complexity SDR solution. However, the performance of SCA degrades as the number of IRS reflectors gets high. Moreover, in [31], the EE maximization problem for an IRS assisted uplink multiple-input-multiple-output (MIMO) NOMA scenario was studied. To achieve their research goals, the authors proposed an iterative approach to solve this multi-variable non-convex problem where they used an SDP iterative approach and difference of convex (DC) programming to optimize the IRS phase shifts. Nevertheless, the complexity of their proposed technique is indeed high as it is based on SDP. Amongst the other works in literature, [32] proposed a hybrid NOMA-OMA multiple access scheme in active IRS aided energy-constrained IoT systems to assist the uplink transmission from multiple IoT devices to an access point. The authors' formulated the problem to maximize the sum throughput by optimizing the IRS beamforming vectors across time and resource allocation. In [33], the authors proposed

using IRS assisted uplink NOMA transmission to reduce the interference in enhanced mobile broadband (eMBB) and ultra-reliable low-latency communications (URLLC) devices. Both [32], [33] have used the high complexity SDP approach to solve the presented problems. Finally, a low complexity design for IRS assisted uplink sparse code multiple access (SCMA) was proposed in [34], however, the presented design considered SCMA not the general uplink power domain NOMA scheme as our proposed system model in this manuscript.

On the other hand, a range of optimization algorithms were also proposed for optimizing IRS-assisted downlink NOMA systems. For instance, the rate performance of IRS-assisted downlink NOMA was optimized in [35] by maximizing the minimum decoding signal-to-interference-plus-noise-ratio (SINR) of all the users by leveraging the alternating optimization and SDR techniques to jointly optimize the NOMA power allocation and IRS phase shifts. Additionally, in [36], the authors solved the sum rate maximization problem of a downlink multiple-input-single-output (MISO) IRS-assisted NOMA system using alternating optimization and SDP-based sequential rank-one relaxation approaches to jointly optimize active beamforming at the BS and passive beamforming at the IRS. In [37], the EE of an IRS-assisted downlink NOMA system was maximized by jointly optimizing the transmit beamforming at the BS and the IRS phase shifts using alternating optimization and SDR. Moreover, the total transmit power minimization at the BS for an IRS-empowered downlink MISO NOMA network was investigated in [38], where the joint active beamforming at the BS and passive beamforming at the IRS problems were formulated as a bi-quadratically constrained optimization problem. Then, the formulated problem was converted to a bi-SDP problem, and a unified DC approach was proposed to obtain a rank-one solution.

It is noteworthy to observe that the formulated optimization problems in most of the previous IRS-NOMA literature have been solved by converting them into SDP problems and then employing either SDR or successive rank-one procedures to obtain the final solution. However, these solutions suffer from extremely high complexity, which makes them impractical for energy-autonomous applications. Moreover, in IRS-based systems, the complexity and time requirement for executing existing optimization algorithms severely increase as the number of reflecting elements increases. Therefore, devising low-complexity, efficient optimization techniques for optimizing IRS-assisted uplink NOMA systems is indispensable for energy-autonomous networks, such as IoT and UAV networks, where the BS, possibly attached to a drone, and users have limited hardware capabilities and energy budgets.

B. Motivation and Contributions

The design of multi-user IRS-assisted uplink NOMA is challenging since it requires the joint optimization of both the transmit powers and the IRS reflection coefficients. The optimization variables are coupled together, which makes the problem intractable for a closed-form solution and highly non-convex to be solved using low-complexity convex optimization algorithms. Therefore, existing solutions for uplink IRS-

¹The assumption that the IRS channel vectors are known is widely accepted as there are specific works on IRS channel estimation [23]–[27].

NOMA design in the literature use SDP approaches to convert the problem into a convex form, which can then be solved using convex programming tools, as mentioned earlier. However, SDP-based approaches have high complexity, especially when the number of IRS elements is high, which renders these solutions unsuitable for IoT applications where the BS may have limited power resources that cannot accommodate very complex tasks.

Motivated by the complexity drawback of existing uplink IRS-NOMA optimization algorithms, we propose a low-complexity and efficient optimization algorithm that suits battery-powered IoT applications with limited processing capabilities. Specifically, we consider the IRS-assisted uplink NOMA system, shown in Fig. 1. Manifold optimization-based algorithms are proposed for the system model introduced in this paper, aiming to provide a low-complexity design with powerful performance. Two main optimization problems, namely, the total users' transmit power minimization and the system's EE maximization, are formulated and solved. It is noteworthy to mention that these two problems form the basis for reducing power consumption or increasing energy efficiency in the network. To solve these problems, we use iterative alternation algorithms to jointly optimize the transmit powers of the users and the phase shifts at the IRS, under QoS and transmit power constraints for the users. In each iteration of the alternating algorithm, the transmit powers optimization is solved given the IRS passive beamforming coefficients; then, the IRS coefficients are optimized given the obtained users' transmit powers. To solve the IRS phase shifts optimization in each iteration, we devise an efficient complex circle manifold (CCM) optimization algorithm. The performance and complexity of the proposed technique are then compared against three SDP-based benchmarks that were used in the literature of IRS-NOMA, which are SDR [6], SDP-difference of convex (SDP-DC) [38], and SROCR [36]. It is noteworthy to highlight that manifold optimization was used in different IRS-assisted communication systems, e.g., [11], [39]–[41]. As for complexity, the proposed CCM technique has a complexity of order $\mathcal{O}(L^2)$, whereas the SDP algorithms have a complexity of order $\mathcal{O}(L^{4.5})$, where L is the number of IRS reflectors.

Despite the benefits of integrating NOMA with IRS in energy-constrained IoT networks, as discussed at the beginning of this subsection, it is necessary to compare the performance gain of integrating IRS-NOMA into energy-constrained IoT systems against the IRS-OMA counterpart. For multi-user IRS-enabled systems, does the IRS-NOMA scheme always outperform the IRS-OMA one in all cases? To answer this question, we proceed to formulate and address both the challenges of minimizing transmit power and maximizing the EE in an uplink OMA system with IRS. This involves the joint optimization of IRS phase shifts and time allocation for each user. The goal is to compare the performance of the IRS-NOMA scheme using the proposed low-complexity CCM-based algorithm against the optimized IRS-OMA counterpart.

In summary, the main contributions of this work are:

- The total transmit power minimization and the EE maximization problems are formulated based on the intro-

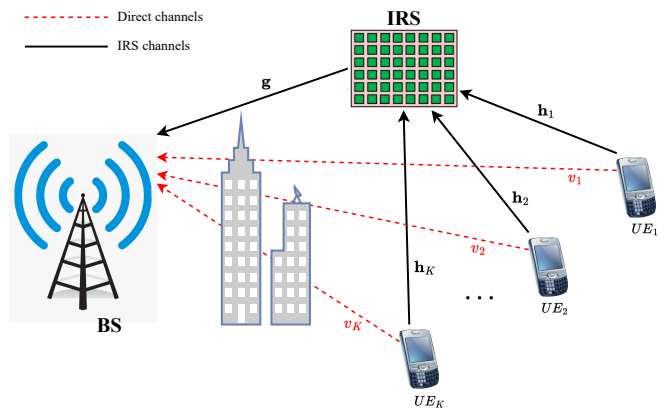


Figure 1: IRS assisted NOMA uplink system model

duced IRS-assisted uplink NOMA system. Accordingly, a joint design for optimizing the users' transmit powers and the IRS phase shifts is then proposed to solve each of the two considered problems using alternating optimization. Moreover, we shed light on the relationship between the two considered optimization problems in the simulation results, which has not been thoroughly investigated in the literature. The results unveil that although the two problems converge to the same solution when the users' uplink target rates are high, they have different solutions when the target rates are low.

- Additionally, to solve the IRS passive beamforming sub-problem, we devise a novel low-complexity manifold optimization-based approach. The proposed manifold optimization-based algorithm can obtain a locally optimal solution that adheres to the unit-modulus constraints of the IRS reflecting coefficients.
- We further formulate and solve both the transmit power minimization and EE maximization problems for the IRS-assisted uplink OMA system, by jointly optimizing the IRS phase shifts and the time allocation for each user, to compare against the IRS-assisted NOMA system. Interestingly, the simulation results reveal that NOMA outperforms OMA when the target user rate constraints are high; however, OMA is preferable when the target uplink rate constraints are low.

The rest of the paper is organized as follows. In Sec. II, the IRS-assisted uplink NOMA system model is presented, while in Sec. III, the total transmit power minimization problem is formulated and solved. The EE maximization problem is presented and solved in Sec. IV. In Sec. V, we discuss the IRS-assisted uplink OMA system. Finally, the numerical results and conclusions are given in sections VI and VII, respectively.

II. SYSTEM MODEL

As shown in Fig. 1, we consider an IRS-assisted uplink NOMA system in which a group of K single-antenna users transmit data symbols to a single-antenna BS in the same time-frequency block. The uplink transmission is assisted by an IRS panel, which consists of L reflecting elements where the i -th element can introduce a phase shift θ_i to the incident signal before reflection. It is worth mentioning that the NOMA

scheme is used in this system model to leverage the variation in the users' distances from the BS, i.e., there exist near users and far users, which is known to be an appropriate scenario in which NOMA can perform well. Using NOMA in this setup is beneficial, in most scenarios, since it can provide more power savings than using OMA by reducing the overall transmit power of the users while satisfying their QoS requirements. The dotted lines in Fig. 1 represent the direct channels between the users and the BS. The direct channels are assumed to be independent and identically distributed (i.i.d.) Rayleigh fading channels because line-of-sight (LoS) paths are not guaranteed in several environments where multiple blockages and obstacles between the users and BS obstruct the communication between them. This actually motivates the deployment of IRS in the considered system model to provide robust links between the users and BS and facilitate communication between them. The direct channel between the k -th user and BS is denoted as v_k . The IRS-BS and the k -th user-IRS channel vectors are denoted as $\mathbf{g} \in \mathbb{C}^{L \times 1}$ and \mathbf{h}_k , respectively, where they are represented by the solid lines in Fig. 1. The IRS panel is typically placed in a position where a LoS path to the BS and to the users always exists in practice. Hence, we adopt Rician fading channels to model the channel vectors \mathbf{g} and \mathbf{h}_k . The channel vector \mathbf{g} is given as

$$\mathbf{g} = \sqrt{\frac{PL(d_{IB})K_{IB}}{K_{IB} + 1}} \mathbf{g}^{LoS} + \sqrt{\frac{PL(d_{IB})}{K_{IB} + 1}} \mathbf{g}^{NLoS}, \quad (1)$$

where K_{IB} denotes the Rician factor of \mathbf{g} , d_{IB} is the distance between the IRS and the BS, and \mathbf{g}^{LoS} and \mathbf{g}^{NLoS} are the line-of-sight (LoS) and non-LoS (NLoS) components, respectively. The LoS component is deterministic, however, the elements of \mathbf{g}^{NLoS} are independent and identically distributed (i.i.d.) complex normal, $\mathcal{CN}(0, 1)$, random variables. The channel vector between user k and the IRS, \mathbf{h}_k , is given as

$$\mathbf{h}_k = \sqrt{\frac{PL(d_{U_k, I})K_{UI}}{K_{UI} + 1}} \mathbf{h}_k^{LoS} + \sqrt{\frac{PL(d_{U_k, I})}{K_{UI} + 1}} \mathbf{h}_k^{NLoS}, \quad (2)$$

where K_{UI} denotes the Rician factor of \mathbf{h}_k , $d_{U_k, I}$ is the distance between user k and the IRS, \mathbf{h}_k^{LoS} and \mathbf{h}_k^{NLoS} are the LoS and NLoS components, respectively. PL represents the path loss which is modeled for all the channels as

$$PL(d) = \eta_0 \left(\frac{d}{d_0} \right)^{-\alpha}, \quad (3)$$

where η_0 is the path loss at the reference distance $d_0 = 1$ m, d represents the link distance between the transmitter and the receiver, and α is the path loss exponent.

In Fig. 1 the K users transmit K data symbols *simultaneously*; hence, the received signal at the BS is written as

$$\mathbf{y} = \sum_{k=1}^K (\mathbf{g}^T \mathbf{W} \mathbf{h}_k + v_k) \sqrt{p_k} s_k + n, \quad (4)$$

where s_k is the data symbol transmitted by user k , $\mathbb{E}[|s_k|^2] = 1$, p_k is the transmitted power from user k , and n is the complex Gaussian $\mathcal{CN}(0, \sigma_n^2)$ additive white Gaussian noise (AWGN) at the BS. \mathbf{W} is an $L \times L$ diagonal matrix, i.e.

$\mathbf{W} = \text{diag}\{\mathbf{w}\}$, whose diagonal elements are the IRS reflection coefficients, where $\mathbf{w} = [e^{j\theta_1}, \dots, e^{j\theta_L}]^T$. The angle $\theta_i \in [0, 2\pi[$ represents the phase shift introduced by i th reflecting element.

In uplink NOMA, the BS performs SIC to successfully decode the K superimposed users' signals. SIC requires ordering of the users according to their effective channel gains. Users with better effective channel gains are decoded first considering the other interfering signals as noise. The combined effective channel of the k th user in this system model is $\mathbf{g}^T \mathbf{W} \mathbf{h}_k + v_k$, which clearly depends on the unknown phase shifts matrix \mathbf{W} . Therefore, in this paper, we order the users according to their maximum achievable effective channel [35]. For example, for user k , we can calculate the maximum achievable effective channel gain at the BS by assuming that the IRS phases are solely optimized to maximize the channel of this user. In this case, the IRS phases should be adjusted to align all the combined IRS channels with the direct link v_k of user k , i.e., $\theta_i = \theta_{v_k} - \theta_{g_i} - \theta_{h_{ki}}$. In this case, the maximum achievable channel gain of user k is calculated as $\sum_{i=1}^L |g_i| |h_{ki}| + |v_k|$. Without loss of generality, we assume that the users are arranged in a descending order as

$$\sum_{i=1}^L |g_i| |h_{1i}| + |v_1| \geq \dots \geq \sum_{i=1}^L |g_i| |h_{Ki}| + |v_K|. \quad (5)$$

Assuming successful decoding of the users at BS using SIC², the SINR of the k th user can be expressed as

$$\gamma_k = \frac{p_k |\mathbf{g}^T \mathbf{W} \mathbf{h}_k + v_k|^2}{\sum_{j=k+1}^K p_j |\mathbf{g}^T \mathbf{W} \mathbf{h}_j + v_j|^2 + \sigma_n^2}, \quad (6)$$

where $\sum_{j=k+1}^K p_j |\mathbf{g}^T \mathbf{W} \mathbf{h}_j + v_j|^2 = 0$ when $k = K$, i.e., the last user in the SIC order suffers no interference. Then, the corresponding achievable data rate of the k th user is given by

$$R_k = \log_2(1 + \gamma_k), \quad \forall k. \quad (7)$$

The main interest of this work is to efficiently optimize the IRS reflection coefficients matrix, \mathbf{W} , and the users' transmit powers, p_k , with the objective of minimize the overall transmit power while satisfying the required data rates. This problem is challenging since the optimization variables, p_k and \mathbf{W} , in (6) are coupled, which makes it difficult to solve at once. The optimization problem and its challenges are presented in Sec. III. Existing solutions for similar optimization problems in the literature rely on SDP approaches, which are computationally extensive, especially when the number of IRS elements is large. Motivated by this, we present an efficient manifold optimization-based approach that provides better performance with much lower computational complexity, as illustrated in Sec. III.

²It is widely accepted to assume the successful decoding of the NOMA users in the capacity equation according to the SIC order [13], [14], [22], [42]–[46]. That is, it is common to assume that decoding can be successful if the user transmits at a rate that does not exceed its capacity limit. Additionally, the NOMA users are assumed to have difference in their distances to the IRS so that their average received signal powers at the BS have a minimum difference that is suitable for employing SIC. The difference in the average received signal power of all the NOMA users is realized by grouping users with appropriate variance in their distances to the IRS.

III. TOTAL TRANSMIT POWER MINIMIZATION

In this section, we aim to minimize the sum of the transmitted powers of all the users in the multi-user IRS assisted NOMA system assuming that each user has a minimum rate requirement that must be guaranteed to satisfy some QoS requirement at each individual user. This minimization is done by jointly optimizing the passive beamforming coefficients at the IRS, \mathbf{W} , and the power allocation at the users. Hence, the sum power minimization problem can be formulated as

$$\min_{\mathbf{p}, \mathbf{W}} \sum_{k=1}^K p_k \quad (8a)$$

$$\text{s.t. } R_k \geq R_k^{\min}, \quad k = 1, 2, \dots, K, \quad (8b)$$

$$p_k \leq P^{\max}, \quad (8c)$$

$$|w_i| = 1, \quad i = 1, 2, \dots, L, \quad (8d)$$

where $w_i = e^{j\theta_i}$ is the (i, i) entry of \mathbf{W} and P^{\max} is the maximum allowable uplink transmit power for the users. Solving the optimization problem in (8) is challenging because the optimization variables, p_k and w_i , are multiplied by each other in constraints (8b), and the constraints in (8d) are non-convex. We use the fact that we have two sets of optimization parameters, p_k 's and w_i 's, coupled together in (8), and propose an alternating optimization algorithm to solve it in \mathbf{p} and \mathbf{W} in an iterative manner.

During each iteration n in our proposed alternating optimization algorithm, we solve the power allocation problem to get \mathbf{p}^n first for a given \mathbf{W}^{n-1} , then we optimize \mathbf{W} for a given \mathbf{p}^n to get \mathbf{W}^n , and so on. Therefore, the original problem (8) is separated into two sub-problems that are solved alternatively, as explained in the next subsections.

A. Optimization of users' transmit powers

In the n th iteration, given the value of \mathbf{W}^n , the power allocation problem in \mathbf{p} reduces to

$$\min_{\mathbf{p}} \sum_{k=1}^K p_k \quad (9a)$$

$$\text{s.t. } C_k(\mathbf{W}^n, \mathbf{p}) = p_k \left| \mathbf{g}^T \mathbf{W}^n \mathbf{h}_k + v_k \right|^2 - (2^{R_k^{\min}} - 1) \\ \times \left(\sum_{j=k+1}^K p_j \left| \mathbf{g}^T \mathbf{W}^n \mathbf{h}_j + v_j \right|^2 + \sigma_n^2 \right) \geq 0, \quad (9b)$$

$$k = 1, 2, \dots, K, \\ p_k \leq P^{\max}, \quad (9c)$$

$$p_k \geq 0. \quad (9d)$$

It is worth noting that there is no need for adding extra constraint for the SIC decoding order since it is already embedded in the rate constraints of the NOMA users (9b). That is, given the reflection coefficient matrix \mathbf{W} , the constraints in (9b) must be satisfied with equality as an optimality condition, which guarantees the correct SIC order.

The optimization problem (9) can be expressed as a linear program in \mathbf{p} in the form of [47]

$$\min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} \quad (10a)$$

$$\text{s.t. } \mathbf{A} \mathbf{x} \leq \mathbf{b} \quad (10b)$$

$$\mathbf{x} \geq \mathbf{0}, \quad (10c)$$

where the constraints in (9b) and (9c) are affine inequalities in \mathbf{p} which can be all written in the form of (10b), and the cost function (9a) is linear having the form of (10a) with \mathbf{c} equals to an all ones vector. Therefore, problem (9) is a linear program in \mathbf{p} which can be easily solved using the standard *simplex method*. The simplex method is a well known algorithm to solve any linear program, having a linear objective function and affine constraints, using slack variables and pivot variables as a means to find the optimal solution [47], [48].

In the next subsection, we present a novel and low complexity approach for solving the challenging optimization problem (8) in \mathbf{W} given the values of p_k 's. We first present an SDP approach as a benchmark from the current solutions in the literature, followed by our proposed novel and low complexity CCM based optimization approach. The performance of the two approaches will be compared along with other current SDP schemes, in the results section.

B. Optimization of the IRS reflection coefficients

Firstly, we present a current SDR benchmark scheme and emphasize the drawbacks of existing counterparts.

For some given p_k values in the n th iteration, the optimization problem in (8) becomes a feasibility-check problem in \mathbf{W} , since the cost function in (8a) does not depend on the IRS coefficients \mathbf{W} . Hence, given the power allocation vector \mathbf{p}^n , the feasibility-check problem in \mathbf{W} is given as

$$\text{Find } \mathbf{W} \quad (11a)$$

$$\text{s.t. } C_k(\mathbf{W}, \mathbf{p}^n) \geq 0, \quad k = 1, 2, \dots, K, \quad (11b)$$

$$|w_i| = 1, \quad i = 1, 2, \dots, L. \quad (11c)$$

1) *SDR benchmark for optimizing \mathbf{W}* : Optimizing \mathbf{W} in problem (11) can be solved using SDR by converting problem (11) into a SDP as follows. Problem (11) is non-convex because of the non-convex constraints in (11b) and the unit modulus constraints in (11c). However, by converting (11) into a SDP, the problem is transformed to a convex problem and solved using standard tools such as CVX. The constraints in (11b) can be converted into quadratic forms w.r.t. the passive beamforming vector \mathbf{w} , then converting them to convex affine constraints using the SDP technique as follows. The squared magnitude of the effective channel of user k can be written as

$$\left| \mathbf{g}^T \mathbf{W} \mathbf{h}_k + v_k \right|^2 = \left| \mathbf{g}^T \mathbf{H}_k \mathbf{w} + v_k \right|^2, \quad (12)$$

where $\mathbf{H}_k = \text{diag}(\mathbf{h}_k)$ and $\mathbf{w} \in \mathbb{C}^{L \times 1}$ is a column vector containing the diagonal elements of \mathbf{W} . Then, we can expand the squared magnitude of the effective channel of the k th user as

$$\left| \mathbf{g}^T \mathbf{H}_k \mathbf{w} + v_k \right|^2 = (\mathbf{w}^H \mathbf{H}_k^H \mathbf{g}^* + v_k^*) (\mathbf{g}^T \mathbf{H}_k \mathbf{w} + v_k) \quad (13)$$

$$= (\mathbf{w}^H \mathbf{z}_k + v_k^*) (\mathbf{z}_k^H \mathbf{w} + v_k), \quad (14)$$

where $\mathbf{z}_k = \mathbf{H}_k^H \mathbf{g}^* \in \mathbb{C}^{L \times 1}$. Hence, the squared effective channel gain of user k can be given as a quadratic term in \mathbf{w} as

$$|\mathbf{g}^T \mathbf{H}_k \mathbf{w} + v_k|^2 = \mathbf{w}^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{w} + \mathbf{w}^H \mathbf{z}_k v_k + \mathbf{z}_k^H \mathbf{w} v_k^* + |v_k|^2. \quad (15)$$

Equation (15) can be expressed as $\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}} + |v_k|^2$, where

$$\bar{\mathbf{w}} = \begin{bmatrix} \mathbf{w} \\ 1 \end{bmatrix} \text{ and } \mathbf{B}_k = \begin{bmatrix} \mathbf{z}_k \mathbf{z}_k^H & \mathbf{z}_k v_k \\ \mathbf{z}_k^H v_k^* & 0 \end{bmatrix}. \quad (16)$$

Knowing that $\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}} = \text{trace}(\bar{\mathbf{w}}^H \mathbf{B}_k \bar{\mathbf{w}}) = \text{trace}(\mathbf{B}_k \bar{\mathbf{w}} \bar{\mathbf{w}}^H)$, and by letting $\widehat{\mathbf{W}} = \bar{\mathbf{w}} \bar{\mathbf{w}}^H \in \mathbb{C}^{(L+1) \times (L+1)}$, we can convert the problem into SDP. The matrix $\widehat{\mathbf{W}}$ is the outer product of the vector $\bar{\mathbf{w}}$ with itself, hence $\widehat{\mathbf{W}}$ is a symmetric positive semi-definite matrix whose rank is 1. Consequently, by introducing the slack variable α to maximize the minimum of the constraints in (11b), problem (11) can be reformulated as

$$\max_{\widehat{\mathbf{W}}, \alpha} \alpha \quad (17a)$$

$$\text{s.t. } p_k \left(\text{trace}\{\mathbf{B}_k \widehat{\mathbf{W}}\} + |v_k|^2 \right) - (2^{R_k^{\min}} - 1) \times \left(\sum_{j=k+1}^K p_j \left(\text{trace}\{\mathbf{B}_j \widehat{\mathbf{W}}\} + |v_j|^2 \right) + \sigma_n^2 \right) - \alpha \geq 0, \forall k, \quad (17b)$$

$$\widehat{\mathbf{W}}(i, i) = 1, \quad i = 1, 2, \dots, L+1, \quad (17c)$$

$$\widehat{\mathbf{W}} \succeq \mathbf{0}, \quad \text{rank}(\widehat{\mathbf{W}}) = 1, \quad (17d)$$

$$\alpha \geq 0. \quad (17e)$$

It should be noted that introducing the slack variable α in (17) is a standard method for maximizing the minimum of some functions [6]. Obviously, problem (17) is a SDP, however, still non-convex because of the non-convex rank one constraint in (17d). By dropping (17d), SDP becomes convex and can be solved using CVX. Assuming the optimal solution of the SDR of (17) is $\widehat{\mathbf{W}}^*$, then we need to convert $\widehat{\mathbf{W}}^*$ to a feasible solution of the original problem (11). The rank of the obtained solution, $\widehat{\mathbf{W}}^*$, of the SDR is generally greater than 1, thus we will always need to reduce our solution to a rank one matrix. An effective way to find a good rank one approximation is to select the eigenvector of $\widehat{\mathbf{W}}^*$ that corresponds to its maximum eigenvalue [49]. Assuming that the maximum eigenvalue of $\widehat{\mathbf{W}}^*$ is λ_1 and the corresponding eigenvector is \mathbf{q}_1 , then $\bar{\mathbf{w}}^* = \sqrt{\lambda_1} \mathbf{q}_1$ can be considered a solution to our original problem (11). However, this solution maybe still infeasible and does not satisfy the unit modulus constraints for w_i 's. To map the solution to a nearby feasible solution, the elements of $\bar{\mathbf{w}}^*$ are normalized to form $\tilde{\mathbf{w}}$ whose elements are defined as $\tilde{w}_i = \frac{\bar{w}_i^*}{|\bar{w}_i^*|}, \forall i \in \{1, \dots, L+1\}$. Finally, the SDR-based solution of the IRS passive beamforming sub-problem is extracted from $\tilde{\mathbf{w}}$ as $\mathbf{w}^* = \frac{\tilde{\mathbf{w}}_{[1:L]}}{\tilde{w}_{L+1}}$, where $\tilde{\mathbf{w}}_{[1:L]}$ contains the first L elements of $\tilde{\mathbf{w}}$, and \tilde{w}_{L+1} is the last element of it.

We can see that the SDP approaches generally broaden the original feasibility region to a larger set. This expansion is achieved by transforming the optimization variable from a vector, $\mathbf{w} \in \mathbb{C}^{L \times 1}$, to an $L \times L$ matrix in $\mathbb{C}^{L \times L}$, denoted as $\mathbf{W} = \mathbf{w} \mathbf{w}^H$, which is the outer product of the original optimization vector. Subsequently, the optimal solution of the

newly expanded SDP problem, denoted as \mathbf{W}^* , undergoes projection onto the original feasible set to derive a solution for the original optimization vector, \mathbf{w}^* . The projection is executed using Cholesky decomposition, which may sometimes yield results deviating from the optimal solution to the original problem. This deviation contributes to a relative reduction in performance when compared to the proposed CCM-based approach.

Moreover, since the SDP based solutions, like the SDR solution in (17), transform the original non-convex problem in (11) to a convex dual by replacing the original optimization vector, \mathbf{w} , with the outer product matrix, $\mathbf{W} = \mathbf{w} \mathbf{w}^H$, the number of the optimization variables of the new problem expands from L to L^2 . Therefore, the complexity of the existing SDP solutions are extremely high since they search over a far larger feasible set with L^2 optimization variables which are the elements of the optimization matrix, \mathbf{W} . However, as illustrated next, the proposed CCM based algorithm searches over the original feasible set, which is the complex circle manifold, for the original optimization vector, \mathbf{w} , with far less number of optimization variables, L . This leads to a substantial complexity reduction in our algorithm compared to the counterparts as shown in Sec. III-D.

2) *The proposed CCM based approach for optimizing \mathbf{W} :* In this subsection, we introduce an efficient and low complexity solution for optimizing the passive beamforming vector in (11), which is the proposed manifold optimization based algorithm. As long as the cost function in (8a) does not depend on \mathbf{W} , we propose to solve (11) in the second step of each iteration of the proposed alternating optimization algorithm to expand the feasibility region of (9) in the first step of the next iteration, i.e., when optimizing \mathbf{p} given \mathbf{W} that has been optimized in a previous iteration. The feasibility region expansion means that the IRS coefficients are optimized to control and expand the boundaries of the feasibility region of the power allocation sub-problem during the next iteration of the alternating optimization algorithm. Expanding the feasibility region of (8) directly requires increasing the achievable users' SINRs in the constraints in (8b), which gives higher tolerance for the power to be reduced in the power allocation sub-problem since the SINRs are already beyond the boundaries. Consequently, the feasibility expansion technique is realized by maximizing the minimum of the SINR constraints, which results in increasing all the users' SINRs. This, in turn, enhances the convergence of finding a solution to (11). The logic of this is that all the constraints are already satisfied with equality by the optimal solution of the sum power minimization sub-problem in (9). Therefore, optimizing \mathbf{W} , to enforce the users' achievable data rates to be larger than the minimum target rates, forthrightly results in the reduction of the total transmit power in (9). Therefore, the problem of maximizing the minimum of the constraints shall be written as

$$\max_{\mathbf{W}} \min\{C_1(\mathbf{W}, \mathbf{p}^n), C_2(\mathbf{W}, \mathbf{p}^n), \dots, C_K(\mathbf{W}, \mathbf{p}^n)\} \quad (18)$$

$$\text{s.t. } (11b), (11c).$$

The min function in (18) is not smooth. Hence, we use a smooth and differentiable approximation as

$$\min_k |C_k| = \left(\sum_{k=1}^K C_k^{-\omega} \right)^{-1/\omega}, \text{ as } \omega \rightarrow +\infty, \quad (19)$$

and $|C_k| = C_k$ because we know that all the constraints are non negative, i.e., $C_k \geq 0$. Therefore, the manifold optimization problem can be written as

$$\begin{aligned} \max_{\mathbf{w}} \quad & \left(\sum_{k=1}^K C_k^{-\omega} \right)^{-1/\omega} \\ \text{s.t.} \quad & (11b), (11c). \end{aligned} \quad (20)$$

The unit-modulus constraints in (11c) restrict the optimization vector, \mathbf{w} , to be located on the surface of a smooth Riemannian manifold contained in \mathbb{C}^L . Precisely, all the optimization variables, w_i , lie on a continuous surface called the CCM, which is defined as

$$\mathcal{S} = \{w_i \in \mathbb{C} : |w_i| = 1\}. \quad (21)$$

The circle, \mathcal{S} , forms a smooth sub-manifold of \mathbb{C} which is a Riemannian manifold. Due to having L optimization variables in (20), the feasible set of the optimization problem is the Cartesian product of L complex circles which is given as

$$\mathcal{S}_1 \times \mathcal{S}_2 \times \dots \times \mathcal{S}_L. \quad (22)$$

The Cartesian product of smooth Riemannian manifolds forms a smooth Riemannian sub-manifold of \mathbb{C}^L . Hence, the feasible set of (20) is an L -dimensional CCM which is formally defined as

$$\begin{aligned} \mathcal{S}^L \triangleq \mathcal{S}_1 \times \dots \times \mathcal{S}_L = \{ \mathbf{w} = [w_1, \dots, w_L] \in \mathbb{C}^L \\ : |w_1| = \dots = |w_L| = 1 \}. \end{aligned} \quad (23)$$

The optimization problem in (20) has extra constraints in (11b) other than the unit-modulus constraints. In the following, we explain how to manage the constraints given in (11b), to solve (20). We propose to handle these additional constraints by using a standard approach called the exact penalty method. To account for the constraints in (11b), the exact penalty method adds a weighted penalty term for each constraint to the the objective function being optimized. When one of the constraints is violated, the modified objective function is largely penalized by moving far away from the optimum point. By using the exact penalty method, (20) is reduced to an unconstrained optimization problem, but generally having a non-smooth objective function. Therefore, in the Riemannian case, the exact penalty method transforms (20) to an unconstrained optimization problem over the L -dimensional complex circle manifold as follows

$$\max_{\mathbf{w} \in \mathcal{M}} \left(\sum_{k=1}^K C_k(\mathbf{w})^{-\omega} \right)^{-1/\omega} + \rho \left(\sum_{k=1}^K \max\{0, -C_k(\mathbf{w})\} \right), \quad (24)$$

where $\rho > 0$ is a penalty weight and \mathcal{M} is the L -dimensional complex circle manifold over which the problem is optimized. Note that the constant modulus constraints in (11c) are satis-

fied by restricting the feasible set to the manifold \mathcal{M} . In the exact penalty method, only a finite penalty weight, ρ , is needed to exactly satisfy the constraints, hence the method's name, [50]. The penalized objective function in (24) is neither smooth nor differentiable due to the existence of the max functions that replace the constraints with the corresponding penalties. Therefore, we can use a smoothing technique to smooth and solve (24). Like the case of the min function in (18), we need to represent the max function in (24) using a smooth and differentiable alternative, so that we can derive the Euclidean gradient and the corresponding Riemannian gradient to be used in the manifold optimization algorithm steps. Without the smooth alternative, the derivation of the Riemannian gradient of the original version in (24) may not be possible. By using the linear-quadratic loss approach [51], the max function in (24) can be approximated, using a smoothing parameter $u > 0$, as $\max\{0, x\} \approx \mathcal{P}(x, u)$, where $\mathcal{P}(x, u)$ is given by

$$\mathcal{P}(x, u) = \begin{cases} 0 & x \leq 0 \\ \frac{x^2}{2u} & 0 \leq x \leq u \\ x - \frac{u}{2} & x \geq u, \end{cases} \quad (25)$$

Hence, a smooth and differentiable unconstrained version of our manifold optimization problem can be written as

$$\max_{\mathbf{w} \in \mathcal{M}} Q(\mathbf{w}) = \left(\sum_{k=1}^K C_k(\mathbf{w})^{-\omega} \right)^{-1/\omega} + \rho \left(\sum_{k=1}^K \mathcal{P}(-C_k(\mathbf{w}), u) \right). \quad (26)$$

Thus, the problem in (20) is converted to a smooth unconstrained manifold optimization problem whose feasible points lie on the surface of the complex circle manifold, \mathcal{S}^L . In that case, we can utilize gradient-based manifold optimization techniques to solve (26).

Similar to the case of Euclidean spaces, there are two main steps for a gradient-descent based algorithm on Riemannian manifolds. Firstly, a descent direction should be found, then a step size is computed along this direction. Afterwards, the solution is updated iteratively until convergence by repeating the two steps in each iteration. However, calculating a descent direction should be adjusted to take into account the geometric nature of the manifold, which is discussed in the following. The tangent space, $T_{\mathbf{w}}\mathcal{M}$, at a point, \mathbf{w} , on a differentiable manifold, \mathcal{M} , is defined as the real vector space that tangentially contains the possible directions in which one can tangentially pass through \mathbf{w} . The tangent space at \mathbf{w} is given by

$$T_{\mathbf{w}}\mathcal{M} = \{ \mathbf{v} \in \mathbb{C}^L : \Re(\mathbf{v} \odot \mathbf{w}^*) = \mathbf{0}_L \}, \quad (27)$$

where $\Re(\cdot)$ denotes the element-wise real-part of the complex vector, and \odot is the Hadamard element-wise multiplication. In the context of manifold optimization, the gradient of a function is called the *Riemannian gradient* which is defined as the direction of the steepest increase of the objective function at a certain point \mathbf{w} , and this direction must lie in the tangent space at the point \mathbf{w} . The Riemannian gradient at a point on the manifold is calculated by first calculating the Euclidean gradient at that point then projecting it onto the tangent space at the point using a projection operator. The orthogonal projection operator of a vector \mathbf{v} onto the tangent

space, $T_{\mathbf{w}}\mathcal{M}$, at point \mathbf{w} on the L -dimensional complex circle manifold is given by [50]

$$P_{T_{\mathbf{w}}\mathcal{M}}(\mathbf{v}) = \mathbf{v} - \Re(\mathbf{v} \odot \mathbf{w}^*) \odot \mathbf{w}. \quad (28)$$

Hence, the Riemannian gradient of our smooth objective function Q in (26) on the manifold can be given as

$$\begin{aligned} \nabla_{\mathcal{M}}Q(\mathbf{w}) &= P_{T_{\mathbf{w}}\mathcal{M}}(\nabla Q(\mathbf{w})) \\ &= \nabla Q(\mathbf{w}) - \Re(\nabla Q(\mathbf{w}) \odot \mathbf{w}^*) \odot \mathbf{w}, \end{aligned} \quad (29)$$

where $\nabla Q(\mathbf{w})$ is the Euclidean gradient at the point \mathbf{w} . In the following, we derive the the Euclidean gradient of Q , which is required to obtain the Riemannian gradient. The Euclidean gradient is given by

$$\nabla Q(\mathbf{w}) = \left[\frac{\partial Q}{\partial w_1} \quad \frac{\partial Q}{\partial w_2} \quad \dots \quad \frac{\partial Q}{\partial w_L} \right]^T, \quad (30)$$

where $\partial Q/\partial w_i = \partial Q/\partial \Re(w_i) + j\partial Q/\partial \Im(w_i)$, and $\Im(\cdot)$ denotes the imaginary part of the complex number. Every partial derivative w.r.t. w_i is calculated as

$$\frac{\partial Q}{\partial w_i} = \frac{\partial f}{\partial w_i} + \rho \sum_{k=1}^K \frac{\partial}{\partial w_i} \mathcal{P}(-C_k(\mathbf{w}), u), \quad (31)$$

where f is the first term of the objective function Q in (26), and $\frac{\partial f}{\partial w_i}$ is computed as

$$\begin{aligned} \frac{\partial f}{\partial w_i} &= \frac{\partial f}{\partial \Re(w_i)} + j \frac{\partial f}{\partial \Im(w_i)} \\ &= \left(\sum_{k=1}^K C_k(\mathbf{w})^{-\gamma} \right)^{-1/\gamma-1} \sum_{k=1}^K C_k(\mathbf{w})^{-\gamma-1} C'_{ki}(\mathbf{w}), \end{aligned} \quad (32)$$

where $C'_{ki}(\mathbf{w})$ is calculated as

$$\begin{aligned} C'_{ki}(\mathbf{w}) &= \frac{\partial C_k(\mathbf{w})}{\partial \Re(w_i)} + j \frac{\partial C_k(\mathbf{w})}{\partial \Im(w_i)} \\ &= p_k \left(2\Re(\mathbf{d}_k^T \mathbf{w})(a_{ki} - jb_{ki}) + 2\Im(\mathbf{d}_k^T \mathbf{w})(b_{ki} + ja_{ki}) \right) \\ &\quad - (2^{R_k^{\min}} - 1) \left\{ \sum_{l=k+1}^K p_l \left(2\Re(\mathbf{d}_k^T \mathbf{w})(a_{li} - jb_{li}) \right. \right. \\ &\quad \left. \left. + 2\Im(\mathbf{d}_k^T \mathbf{w})(b_{li} + ja_{li}) \right) \right\}, \end{aligned} \quad (33)$$

where $a_{ki} = \Re(h_{ki}g_i)$, $b_{ki} = \Im(h_{ki}g_i)$ and the vector $\mathbf{d}_k = \mathbf{h}_k \odot \mathbf{g}$ is the Hadamard product of the two vectors \mathbf{h}_k and \mathbf{g} . The partial derivative, $\frac{\partial \mathcal{P}(-C_k(\mathbf{w}), u)}{\partial w_i}$, is calculated as

$$\frac{\partial \mathcal{P}(-C_k(\mathbf{w}), u)}{\partial w_i} = \begin{cases} 0 & -C_k(\mathbf{w}, u) \leq 0 \\ \frac{C_k(\mathbf{w})}{u} C'_{ki}(\mathbf{w}) & 0 \leq -C_k(\mathbf{w}, u) \leq u \\ -C'_{ki}(\mathbf{w}) & -C_k(\mathbf{w}, u) \geq u, \end{cases} \quad (34)$$

where $C'_{ki}(\mathbf{w})$ is calculated in (33). Then, by substituting (32) and (34) in (31), we obtain the Euclidean gradient which is required to calculate the Riemannian gradient in (29) for our algorithm. Algorithm 1 illustrates the steps of solving the unconstrained manifold optimization problem in (26) by updating the penalty coefficient ρ and the smoothing parameter u in an iterative manner.

Algorithm 1: Exact penalty method via smoothing

```

1 Input: Starting point  $\mathbf{w}_0$ , starting penalty coefficient
    $\rho_0$ , starting smoothing accuracy  $u_0$ , minimum
   smoothing accuracy  $u_{\min}$ , constants  $\theta_u \in (0, 1)$ ,
    $\theta_\rho > 1$ ,  $\tau \geq 0$ , minimum step length  $d_{\min}$ .
2 for  $l = 0, 1, 2, \dots$  do
3   To obtain  $\mathbf{w}_{l+1}$ , choose any sub-solver to
   approximately solve
   
$$\min_{\mathbf{w} \in \mathcal{M}} Q(\mathbf{w}, \rho_l, u_l)$$

   with warm-start at  $\mathbf{w}_l$  and stopping criterion
   
$$\|\text{grad } Q(\mathbf{w}, \rho_l, u_l)\| \leq \delta.$$

4   if
    $(\text{dist}(\mathbf{w}_l, \mathbf{w}_{l+1}) < d_{\min} \text{ or } u_l \leq u_{\min})$  and  $C_k(\mathbf{w}_{l+1}) < \tau$ 
   then
5     | Return  $\mathbf{w}_{l+1}$ ;
6   end
7    $u_{l+1} = \max\{u_{\min}, \theta_u u_l\}$ ;
8   if  $(l = 0 \text{ or } -C_k(\mathbf{w}_{l+1}) \geq \tau)$  then
9     |  $\rho_{l+1} = \theta_\rho \rho_l$ 
10  else
11    |  $\rho_{l+1} = \rho_l$ ;
12  end
13 end

```

The feasibility of (20) is ensured by restricting the feasible set of the optimization problem to the complex circle manifold to satisfy the unit modulus constraints in (11c). The other constraints in (11b) are satisfied by penalizing the original cost function as in (24) using the exact penalty method.

Next, we investigate the behavior of Algorithm 1 and explain how the parameters ρ and u are updated in each iteration until convergence. First, we need to mention that the optimum points of the penalized unconstrained problem in (26) are the same as the optimum points of the original problem in (20) when the penalty coefficient ρ is higher than a certain threshold [50]. This threshold is usually unknown; hence, we adopt a common iterative approach provided in [52] to address this problem. This approach does not start with a large ρ because this may slow down the convergence of Algorithm 1. Therefore, Algorithm 1 starts with a relatively low initial value of ρ_0 , then this value is increased in each iteration, by multiplying it by the constant $\theta_\rho > 1$, if the constraints in (11b) are violated as shown in lines 8 and 9 in Algorithm 1. The parameter τ is a tolerance factor which is a low positive number so that if $-C_k(\mathbf{w}_{l+1})$ surpasses τ , the point \mathbf{w}_{l+1} is considered out of the feasible region and ρ must be increased.

The approximation function in (34) is more accurate when the smoothing parameter u_l is decreased. However, if u_l is too small, numerical difficulties may arise when using the approximation function in (34). Consequently, Algorithm 1 starts with u_0 , then keeps decreasing u_l , by multiplying it with the constant fraction θ_u , in each iteration, as in line 7, until it reaches a minimum value u_{\min} . Algorithm 1 terminates when

the Euclidean distance between \mathbf{w}_{l+1} and \mathbf{w}_l is smaller than d_{\min} . To solve the unconstrained optimization problem in line 3 in each iteration, we use the trust region manifold optimization solver [50], that exists in *Manopt* MATLAB toolbox [53], by setting a stopping criterion on the gradient norm.

C. Convergence of the alternating optimization discussion

The complete procedure of the proposed CCM-based alternating optimization algorithm is summarized in Algorithm 2, where we alternately solve (8) by solving the power allocation problem (9) and the passive beamforming problem at IRS (26) using the proposed CCM approach. In each iteration of Algorithm 2, we obtain a solution that is used as the initial point for the next iteration. Specifically, the alternating optimization algorithm starts by solving problem (9) given the beamforming vector, \mathbf{w}_{n-1} , rather than solving (26) given \mathbf{p}_{n-1} . This is intentionally designed because (9) is always feasible in \mathbf{p} for any beamforming vector \mathbf{w} , whereas (26) may not always be feasible for any arbitrary \mathbf{p} . Intuitively, when the obtained solution from (26) achieves a higher user uplink rate than the target minimum rate constraint for user k , the transmit power of user k can be suitably decreased while maintaining all the rate constraints. Consequently, the total transmit power in (9) can be reduced accordingly.

More precisely, we can guarantee the convergence of the proposed alternating optimization solution for (8) using the following argument. We can deduce that the cost function of (9), which is the sum of transmit powers at the users, always decreases over the iterations when solving (9) and (26) iteratively using Algorithm 2. This proposition can be proven as follows. When solving (26) in the n -th iteration given the transmit power vector \mathbf{p}_n , the pair $(\mathbf{p}_n, \mathbf{w}_n)$ must be feasible and satisfy the constraints of the original problem (8). This is simply because both the sub-problems (9) and (26) have the same constraints as (8). However, the obtained \mathbf{w}_n expands the feasibility region of (9) in the $(n+1)$ -th iteration of Algorithm 2, which directly leads to a solution of \mathbf{p}_{n+1} that corresponds to a lower cost function of (9), and so on. The algorithm keeps decreasing the cost function of (9) over the iterations until, during some iteration, problem (26) cannot further expand the feasibility region by increasing the achievable SINR of the users. Therefore, (26) will continue to yield the same solution for \mathbf{w} as the previous iteration, which causes Algorithm 2 to converge at such an iteration with no further reduction in the cost function of (9).

In other words, assume that the objective function of (9), based on a certain IRS coefficients vector \mathbf{w} , is denoted as $f(\mathbf{w}, \mathbf{p})$. In the n -th iteration of Algorithm 2, if a feasible solution to (26) exists, i.e., $(\mathbf{w}^n, \mathbf{p}^n)$, then this solution is also feasible for problem (9). Then, it follows that $f(\mathbf{w}^n, \mathbf{p}^{n+1}) \stackrel{(a)}{\leq} f(\mathbf{w}^n, \mathbf{p}^n) \stackrel{(b)}{=} f(\mathbf{w}^{n-1}, \mathbf{p}^n)$; (a) follows because, for a given \mathbf{w}^n , \mathbf{p}^{n+1} is the optimal solution to problem (9). Moreover, (b) follows since the objective function of (9) does not depend on \mathbf{w} and is only affected by \mathbf{p} .

Algorithm 2: Sum power minimization alternation optimization algorithm

```

1 Initialize iteration number:  $n = 0$ .
2 Initial feasible point:  $\mathbf{p}_0, \mathbf{W}_0$ .
3 while  $\Xi \geq \epsilon$  do
4   Solve (9) in  $\mathbf{p}$  given fixed IRS coefficients matrix
    $\mathbf{W}_{n-1}$ ;
5   Solve (26) in  $\mathbf{W}$ , using Algorithm 1, given fixed
   users' power allocation vector  $\mathbf{p}_n$ ;
6   Calculate  $\Xi = \sum_{k=1}^K p_{k,n-1} - \sum_{k=1}^K p_{k,n}$ ;
7    $n \rightarrow n + 1$ ;
8 end
9 Return:  $\mathbf{p}^*, \mathbf{W}^*$ .

```

D. Complexity analysis

In each iteration of the alternating optimization algorithm, we solve the passive beamforming optimization at the IRS either with the proposed manifold optimization-based method in Algorithm 1 or with the SDR benchmark. The complexity of the SDR sub-problem in (17) is determined by the complexity of the *interior point method* used for solving the semidefinite program, which is given as $\mathcal{O}(\max(L+1, K)^4(L+1)^{1/2} \log(1/\zeta))$ [54], where ζ is the accuracy of the solution.

Now, we then analyze the complexity of the proposed manifold optimization-based method presented in Algorithm 1. The complexity of Algorithm 1 is determined by the complexity of using the *trust region method* [53] to solve the unconstrained manifold optimization problem in line 3 of each iteration of the algorithm. The complexity of solving the manifold optimization problem mainly depends on calculating the Euclidean gradient of the cost function, which is given as $\mathcal{O}(L^2)$ [39], [40]. As long as our optimization problem has K constraints for the users, the complexity of calculating the Euclidean gradient of Q in (30) can be given as $\mathcal{O}(KL^2)$. Therefore, the overall complexity of Algorithm 1 can be written as $\mathcal{O}(\Omega_{Alg1}KL^2)$, where Ω_{Alg1} is the number of iterations of Algorithm 1.

From the above discussion, we can easily deduce that the proposed manifold optimization-based algorithm has much lower complexity, i.e., order 2, with respect to L compared to the SDR benchmark, whose complexity is of order 4.5.

IV. ENERGY EFFICIENCY MAXIMIZATION

In this section, EE maximization is discussed for the uplink NOMA scenario in Fig. 1. The EE of our considered system model is defined as the ratio of the achievable sum-rate of the K users in the NOMA resource block over the total sum of the consumed transmission power by the users. Our objective in this section is to maximize the EE while guaranteeing the minimum QoS constraint at each user, i.e., $R_k \geq R_k^{\min}$. The

achievable sum rate of the NOMA scheme can be rewritten as

$$\begin{aligned} R^{sum} &= \sum_{k=1}^K \log_2 \left(1 + \frac{p_k |\mathbf{g}^T \mathbf{W} \mathbf{h}_k + v_k|^2}{\sum_{j=k+1}^K p_j |\mathbf{g}^T \mathbf{W} \mathbf{h}_j + v_j|^2 + \sigma_n^2} \right) \\ &= \log_2 \left(1 + \frac{\sum_{k=1}^K p_k |\mathbf{g}^T \mathbf{W} \mathbf{h}_k + v_k|^2}{\sigma_n^2} \right). \end{aligned} \quad (35)$$

In this section, we adopt the practical power consumption model used in [7] to model the total power dissipated in our IRS-NOMA system. The considered total power consumption model can be expressed as [7]

$$P_{\text{total}} = \sum_{k=1}^K (\nu_k^{-1} p_k + P_{\text{UE},k}) + P_{\text{BS}} + L P_l, \quad (36)$$

where ν_k is the efficiency of the transmit power amplifiers of the users' devices, $P_{\text{UE},k}$ is the hardware static power consumption at the users, P_{BS} is the hardware static power dissipation at BS, and P_l denotes the power consumption of each reflection element of the IRS. Therefore, our considered EE maximization is formulated as

$$\max_{\mathbf{p}, \mathbf{w}} \frac{\log_2 \left(1 + \frac{\sum_{k=1}^K p_k |\mathbf{g}^T \mathbf{H}_k \mathbf{w} + v_k|^2}{\sigma_n^2} \right)}{P_{\text{total}}} \quad (37a)$$

$$\text{s.t. } p_k \leq P^{\max}, \quad (37b)$$

$$C_k(\mathbf{w}, \mathbf{p}) \geq 0, \quad k = 1, 2, \dots, K, \quad (37c)$$

$$|w_i| = 1, \quad i = 1, 2, \dots, L, \quad (37d)$$

where $\mathbf{H}_k = \text{diag}(\mathbf{h}_k)$, P^{\max} is the maximum transmit power allowed at the users, and \mathbf{w} is a column vector that contains the diagonal elements of \mathbf{W} . We propose to solve (37) using alternating optimization as in the previous section because we have two blocks of optimization parameters \mathbf{p} and \mathbf{w} . In the next subsections, we first solve the problem in \mathbf{p} given \mathbf{w} , then the problem is solved in \mathbf{w} given \mathbf{p} .

A. Optimizing the transmit power vector

Given the beamforming vector \mathbf{w} , the optimizing (37) is a non-linear fractional program in \mathbf{p} which can be expressed as

$$\max_{\mathbf{p}} \frac{\log_2 \left(1 + \frac{\sum_{k=1}^K p_k a_k}{\sigma_n^2} \right)}{P_{\text{total}}} \quad (38a)$$

$$\text{s.t. } p_k \leq P^{\max}, \quad (38b)$$

$$C_k(\mathbf{p}; \mathbf{w}) \geq 0, \quad k = 1, 2, \dots, K, \quad (38c)$$

where $a_k = |\mathbf{g}^T \mathbf{H}_k \mathbf{w} + v_k|^2$. The feasible set of (38) is convex since all the constraints are affine in \mathbf{p} . The fractional objective function has a concave numerator w.r.t. p_k , $k \in \{1, \dots, K\}$, while the denominator is an affine mapping over the feasible set of \mathbf{p} . Therefore, this non-linear fractional program is *pseudo-concave* and has an optimal solution that can be efficiently obtained using the iterative Dinkelbach's algorithm [55]. The specific procedure of this algorithm is summarized in Algorithm 3. During each iteration, the proposed algorithm

needs to solve the following problem

$$\begin{aligned} \max_{\mathbf{p}} \quad & \log_2 \left(1 + \frac{\sum_{k=1}^K p_k a_k}{\sigma_n^2} \right) - \beta (P_{\text{total}}) \\ \text{s.t.} \quad & (38b), (38c), \end{aligned} \quad (39)$$

where β is a known constant. Obviously, (39) is concave and its feasible set is convex, hence it can be solved using standard optimization algorithms like the interior point method. Nonetheless, (39) has a specific structure that can be exploited to develop a low complexity solution. The standard approaches are computationally intensive especially that solving (39) is repeated in each iteration of Algorithm 3.

Algorithm 3: EE max. for \mathbf{p}

```

1 Initialize parameters:  $\epsilon > 0, \beta = 0, F > \epsilon.$ 
2 while  $F > \epsilon$  do
3    $\mathbf{p}^* = \arg \max \log_2 \left( 1 + \frac{\sum_{k=1}^K p_k a_k}{\sigma_n^2} \right) - \beta (P_{\text{total}});$ 
   s.t. (38b), (38c);
4    $F =$ 
    $\log_2 \left( 1 + \frac{\sum_{k=1}^K p_k^* a_k}{\sigma_n^2} \right) - \beta \left( \sum_{k=1}^K \nu_k^{-1} p_k^* + c \right),$ 
   where  $c = \sum_{k=1}^K P_{\text{UE},k} + P_{\text{BS}} + L P_l;$ 
5    $\beta = \frac{\log_2 \left( 1 + \frac{\sum_{k=1}^K p_k^* a_k}{\sigma_n^2} \right)}{\sum_{k=1}^K \nu_k^{-1} p_k^* + c};$ 
6 end
```

In the following, we propose a low-complexity optimal solution for (39) to reduce the computational load since Algorithm 3 requires solving (39) many times. Let $F = \log_2 \left(1 + \frac{\sum_{k=1}^K p_k a_k}{\sigma_n^2} \right) - \beta (P_{\text{total}})$, then the partial derivative of F w.r.t. p_k is calculated as

$$\frac{\partial F}{\partial p_k} = \frac{a_k}{(\ln 2)(\sum_{i=1}^K p_i a_i + \sigma_n^2)} - \beta \nu_k^{-1}. \quad (40)$$

By setting $\partial F / \partial p_k = 0$, we get

$$p_k^* = 1 / (\beta \nu_k^{-1} \ln 2) - \left(\sum_{i \neq k} p_i a_i + \sigma_n^2 \right) / a_k. \quad (41)$$

We calculate p_k^* assuming all the remaining $p_i \forall i \neq k$ values are fixed, and this process is repeated until convergence. On this basis, the proposed low-complexity algorithm goes as follows: we first allocate the minimum required power for each user; then, we update the power for the users one by one as illustrated in Algorithm 4; this update continues until convergence. Note that the convergence is guaranteed since F increases or remains unchanged after each update, and F has an upper bound. Moreover, the obtained local optimum is also the global optimum since the cost function F in (39) is concave. p_k^{\min} and p_k^{\max} in Algorithm 4 are calculated for each user according to its boundaries as in the constraints in (38c).

B. Optimizing the IRS coefficients vector \mathbf{w} given \mathbf{p}

In this subsection, we discuss the step of optimizing the IRS coefficients vector, \mathbf{w} , given the transmit power vector \mathbf{p} for the sake of maximizing the EE of the NOMA system

Algorithm 4: Low complexity and fast iterative solution for solving step 3 in Algorithm 3

```

1 Initialize: Set  $p_k = p_k^{\min}$ ,  $k = 1, 2, \dots, K$ ;
2 while 1 do
3    $\mathbf{p}_{old} = \mathbf{p}$ ;
4   for  $k = 1:K$  do
5      $p_k^* = 1/(\beta\nu_k^{-1} \ln 2) - (\sum_{i \neq k} p_i a_i + \sigma_n^2)/a_k$ ;
6     if  $p_k^* < p_k^{\min}$  then
7        $p_k = p_k^{\min}$ ;
8     end
9     if  $p_k^* > p_k^{\max}$  then
10       $p_k = p_k^{\max}$ ;
11    end
12    if  $p_k^{\min} < p_k^* < p_k^{\max}$  then
13       $p_k = p_k^*$ ;
14    end
15  end
16  if  $|\mathbf{p}_{old} - \mathbf{p}| < 10^{-10}$  then
17    break while loop;
18  end
19 end

```

using the alternating maximization algorithm. Both the SDR and the manifold optimization-based solutions are presented in this subsection like the case of the total power minimization problem in Sec. III. Unlike the total transmit power minimization problem, in the EE maximization problem, we have a direct objective function of \mathbf{w} to be maximized when optimizing the coefficients vector \mathbf{w} for a given \mathbf{p} . When maximizing the objective function in (37a) with respect to \mathbf{w} , the denominator is neglected because it is considered a constant as it is not a function of \mathbf{w} . Hence, we only focus on the numerator, which is the log function. Moreover, the log function is monotonically increasing, so we can only focus on maximizing its argument. Therefore, when considering the SDR optimization scheme using the same mathematical steps as in Sec. III-B1, the NOMA EE maximization problem in \mathbf{w} can be written as

$$\max_{\widehat{\mathbf{W}}} \sum_{k=1}^K p_k (\text{trace}\{\mathbf{B}_k \widehat{\mathbf{W}}\} + |v_k|^2) \quad (42a)$$

$$\text{s.t.} \quad p_k (\text{trace}\{\mathbf{B}_k \widehat{\mathbf{W}}\} + |v_k|^2) - (2^{R_k^{\min}} - 1) \times \left[\sum_{j=k+1}^K p_j (\text{trace}\{\mathbf{B}_j \widehat{\mathbf{W}}\} + |v_j|^2) + \sigma_n^2 \right] \geq 0, \forall k, \quad (42b)$$

(17c), (17d),

where $\widehat{\mathbf{W}} = \overline{\mathbf{w}}\overline{\mathbf{w}}^H$ and $\overline{\mathbf{w}}$ is defined in (16). Similar to the SDR problem in (17), the problem in (42) is an SDP which is non-convex due to the non-convex rank one constraint in (17d). By dropping (17d), (42) becomes clearly convex which can be directly solved using optimization tools such as CVX.

We can also use the manifold optimization scheme to optimize the IRS coefficients vector, \mathbf{w} , given \mathbf{p} when maximizing EE. As discussed in Sec. III-B2, the unit modulus constraints on the elements of \mathbf{w} restrict the feasible region to the complex circle manifold \mathcal{M} . Hence, the EE manifold optimization

problem can be written as

$$\max_{\mathbf{w} \in \mathcal{M}} \sum_{k=1}^K p_k |\mathbf{g}^T \mathbf{H}_k \mathbf{w} + v_k|^2 \quad (43a)$$

$$\text{s.t.} \quad C_k(\mathbf{w}) \geq 0, \quad k = 1, 2, \dots, K. \quad (43b)$$

Following the same procedures as in Sec. III-B2, the problem in (43) is transformed to an unconstrained manifold optimization problem using the exact penalty method. Then, we shall use the same smoothing technique as discussed in Sec. III-B2 to smooth the unconstrained penalized optimization problem. Therefore, the smooth unconstrained NOMA EE manifold optimization problem, in \mathbf{w} given \mathbf{p} , can be written as

$$\max_{\mathbf{w} \in \mathcal{M}} \sum_{k=1}^K p_k |\mathbf{g}^T \mathbf{H}_k \mathbf{w} + v_k|^2 + \rho \left(\sum_{k=1}^K \mathcal{P}(-C_k(\mathbf{w}), u) \right), \quad (44)$$

where \mathcal{P} is the smoothing function defined in (25). Now, the unconstrained problem in (44) is smooth and differentiable which can be solved using the manifold optimization toolbox as discussed in Sec. III-B2. The parameters ρ and u in (44) are updated in the same way illustrated in Algorithm 1.

Surprisingly, as shown in the simulation results in Sec. VI, we found that focusing on expanding the feasible set of the problem in \mathbf{p} by finding the beamforming coefficients vector \mathbf{w} that maximizes the minimum of the constraints $C_k(\mathbf{w})$ is more suitable for the EE maximization problem than maximizing the EE objective function itself. This can be done by the alternation between solving the problem (38) in \mathbf{p} given \mathbf{w} , then solving the problem (26) in \mathbf{w} given \mathbf{p} in an iterative manner. This approach is counter-intuitive as it optimizes the IRS beamforming vector during the \mathbf{w} sub-problem to expand the feasible set of the power allocation problem in (38) during the \mathbf{p} sub-problem, rather than maximizing the EE itself, which is the target of the problem. However, the presented simulation results in Sec. VI show that this approach gives way better EE performance than focusing on maximizing the objective function discussed earlier in this section. The reason behind this behavior is that we optimize the passive beamforming vector, \mathbf{w} , to expand the feasible region over which the power allocation problem in (38) is optimized during the alternating optimization algorithm. Therefore, when maximizing the EE objective function in the power allocation phase in (38), Algorithm 4 can search over a larger feasible set to maximize the EE of the system, which may increase the value of the obtained EE in the next iteration of the alternation maximization algorithm. The procedures of the proposed CCM-based alternation optimization algorithm for the EE maximization problem are summarized in Algorithm 5.

V. COMPARISON WITH OPTIMIZED IRS-ASSISTED OMA

In this section, we discuss IRS-assisted uplink OMA. Since IRS-assisted OMA is much simpler than the NOMA counterpart in terms of optimizing the IRS phases, for fairness' sake, we compare our proposed IRS-NOMA scheme against IRS-OMA to quantify the potential performance gains. Moreover, we examine if the performance of the IRS-OMA scheme can

Algorithm 5: EE maximization alternation optimization algorithm

```

1 Initialize iteration number:  $n = 0$ .
2 Initial feasible point:  $\mathbf{p}_0, \mathbf{W}_0$ .
3 while  $\Xi \geq \epsilon$  do
4   Solve (39) in  $\mathbf{p}$ , using Algorithm 4, given fixed
   IRS coefficients matrix  $\mathbf{W}_{n-1}$ ;
5   Solve (26) or (44) in  $\mathbf{W}$ , using Algorithm 1, given
   fixed users' power allocation vector  $\mathbf{p}_n$ ;
6   Calculate  $\Xi = EE(\mathbf{p}_n, \mathbf{W}_n) - EE(\mathbf{p}_{n-1}, \mathbf{W}_{n-1})$ ,
   where  $EE(\mathbf{p}, \mathbf{W})$  is defined in (37a);
7    $n \rightarrow n + 1$ ;
8 end
9 Return:  $\mathbf{p}^*, \mathbf{W}^*$ .

```

outperform the IRS-NOMA one in any specific cases so we can make a useful conclusion on when one of them is favored over the other. To this end, we study the average transmit power minimization and EE maximization problems for the IRS-OMA scheme, as we did in the NOMA scheme, and provide an optimal solution for both problems.

A. IRS-assisted OMA power minimization

In this subsection, we present an optimal solution for minimizing the average transmit power of OMA users when assisted by IRS. The sum transmit power is minimized while having minimum uplink rate constraints for the users. In OMA, each user takes a fraction of the time frame to transmit its data. Unlike NOMA, in IRS-assisted OMA, each user transmits its data separately within its allocated time fraction, α_k , while other users remain idle. The time fraction coefficients, α_k , are optimized along with the transmit powers, p_k , to minimize the average transmit power of the users. This optimization problem can be expressed as

$$\min_{\alpha_k, p_k} \sum_{k=1}^K \alpha_k p_k \quad (45a)$$

$$\text{s.t. } p_k \leq P^{\max}, \quad (45b)$$

$$\alpha_k \log_2 \left(1 + \frac{p_k c_k}{\sigma_n^2} \right) \geq R_k^{\min}, \quad k = 1, 2, \dots, K, \quad (45c)$$

$$\alpha_k \geq 0, \quad (45d)$$

$$\sum_{k=1}^K \alpha_k = 1, \quad (45e)$$

where $c_k = |\mathbf{g}^T \mathbf{W}_k \mathbf{h}_k + v_k|^2$ and P^{\max} is the maximum allowable uplink transmit power for the users. As long as the users send their data separately during their allocated time, the IRS adjusts its phase shifts, \mathbf{W}_k , independently of each user according to the corresponding channel vector, \mathbf{h}_k , as each user experiences an *interference-free* channel. The IRS phase shifts are chosen so that all the elements of the cascaded user-IRS-BS channel, $h_{ki} w_i g_i \forall i$, are aligned to have the same phase which is the phase of the direct channel v_k , i.e., $\theta_i = \theta_{v_k} - \theta_{g_i} - \theta_{h_{ki}}$. To minimize the average transmit power

of the OMA users, the minimum rate constraints in (45c) is satisfied with equality to reduce p_k as much as possible. When the constraints in (45c) are satisfied with equality, the power of each user p_k can be expressed in a closed-form as

$$p_k = \frac{\sigma_n^2}{c_k} \left(2^{R_k^{\min}/\alpha_k} - 1 \right). \quad (46)$$

By eliminating p_k from (45a), (45) can be reduced to

$$\min_{\alpha_k} \sum_{k=1}^K \frac{\sigma_n^2 \alpha_k}{c_k} \left(2^{R_k^{\min}/\alpha_k} - 1 \right) \quad (47a)$$

$$\text{s.t. } (45b), (45d), (45e). \quad (47b)$$

The feasible set of the optimization problem (47) is clearly convex since the constraints (45d) and (45e) are affine in α_k . Now, we need to check the convexity of the objective function in (47). The objective function is separable in α_k 's, i.e. it can be written as a sum of separate terms in $\alpha_1, \alpha_2, \dots, \alpha_K$. Therefore, the Hessian matrix of (47) is diagonal as $\frac{\partial^2 f_{oma}}{\partial \alpha_k \partial \alpha_j} = 0, k \neq j$, where f_{oma} is the objective function in (47). Since the Hessian matrix is diagonal, its eigenvalues are the diagonal elements themselves, which can be calculated as

$$\frac{\partial^2 f_{oma}}{\partial \alpha_k^2} = \frac{\sigma_n^2 ((\ln 2) R_k^{\min})^2}{c_k \alpha_k^3} 2^{(R_k^{\min}/\alpha_k)}, \quad (48)$$

where $\frac{\partial^2 f_{oma}}{\partial \alpha_k^2}$ is the second partial derivative of f_{oma} w.r.t. α_k . Clearly, the second derivatives in (48) are always positive over the feasible set of (47) since α_k 's are non-negative. Consequently, (47) is strictly convex since it has a convex feasible set as well as a convex objective function, then it can be efficiently solved using CVX or MATLAB's `fmincon` function in the optimization toolbox.

B. IRS-assisted OMA EE maximization

In this subsection, we present the EE maximization problem of the OMA system and propose an optimal solution for it. The EE maximization problem is solved under minimum uplink data rate constraints for the users. The EE maximization problem of the OMA system can be formulated as

$$\max_{\alpha_k, p_k} \frac{\sum_{k=1}^K \alpha_k \log_2 \left(1 + \frac{p_k c_k}{\sigma_n^2} \right)}{\sum_{k=1}^K \alpha_k p_k} \quad (49)$$

$$\text{s.t. } (45b), (45c), (45d), (45e).$$

Assuming that c_1 is greater than the values of c_k of other users, then we focus on allocating all the excess power to user 1, i.e., increasing p_1 , while the powers of the other users are chosen to satisfy their minimum rate constraints with equality. Therefore, the EE maximization problem reduces to

$$\max_{\alpha_k, p_1} \frac{\alpha_1 \log_2 \left(1 + \frac{p_1 c_1}{\sigma_n^2} \right) + \sum_{k=2}^K R_k^{\min}}{\alpha_1 p_1 + \sum_{k=2}^K \alpha_k \left[2^{R_k^{\min}/\alpha_k} - 1 \right] \frac{\sigma_n^2}{c_k}} \quad (50a)$$

$$\text{s.t. } \alpha_1 \log_2 \left(1 + \frac{p_1 c_1}{\sigma_n^2} \right) \geq R_1^{\min}, \quad (50b)$$

(45b), (45d), (45e).

Problem (50) is solved using alternating maximization over p_1 and α_k iteratively until convergence. It can be readily proved that the problem in (50) is concave w.r.t. p_1 . However, (50) is quasi-concave in α_k since the numerator is affine and the denominator is convex w.r.t. α_k as proved in the previous subsection. Therefore, the Dinkelbach algorithm discussed in Sec. IV-A can be used to solve (50) w.r.t. α_k .

VI. SIMULATION RESULTS

In this section, we provide extensive simulation results to validate the proposed optimization algorithms for IRS-assisted uplink NOMA systems. The proposed solutions are compared against the IRS-assisted OMA under different scenarios. For each optimization problem, we compare the performance of the proposed manifold algorithm against three SDP-based benchmarks, which are SDR, SDP-DC, and SROCR. Under each scenario, we present the results in terms of the total transmit power, EE, and the achievable uplink sum rate of the proposed solution. Comparing these three system performance measures of the total transmit power minimization problem against those of the EE maximization problem reveals very useful insights, which relate the two problems. The two problems are solved and compared against each other under two scenarios: the first is when low minimum target uplink rates are required for the users, and the second is when high minimum rate constraints are required for the users. The number of users, K , in the NOMA cluster is assumed to be 3.

For fair comparison, similar simulation parameters to those used in [6] and [36] have also been considered in this section, which can be summarized as follows. The distance between the BS and IRS is assumed to be $d_{IB} = 75$ m. The distances between the 3 users and the IRS are 10, 20, 40 m, whereas the distances between the users and the BS are 30, 50, 200 m. The path loss at the reference distance in (3) is $\eta_0 = 10^{-3}$, and the path loss exponents for the direct links (users to BS), IRS to users links, and BS to IRS link are assumed in this simulation setup as $\alpha_{BU} = 5.5$, $\alpha_{IU} = 2.2$, and $\alpha_{BI} = 2.2$, respectively. The noise power σ_n^2 is set to be -114 dBm, and the Rician factors in (1) and (2) are set to be $K_{IB} = K_{UI} = 2.2$. The maximum transmit power constraint, P^{\max} , in (37b) is set to 10 dBm in the case of low minimum rate constraints, and 25 dBm in the case of high minimum rate constraints for the users.

A. NOMA vs. OMA for low minimum rate constraints

In Fig. 2, we present the total transmission power of the users versus the number of reflecting elements at IRS for different schemes. Low QoS constraints at the users are assumed in Fig. 2 where the minimum target achievable rates at the users are set at 0.2 bits/sec/Hz. We compare the performance of the proposed scheme against SDR, SDP-DC, SROCR, and the IRS assisted OMA scheme. The results of the IRS-OMA scenario are included when optimal time allocation (OTA) and uniform time allocation (UTA) are adopted. The figure shows that our proposed optimization scheme outperforms SDR, SDP-DC, and SROCR, although the complexity of the

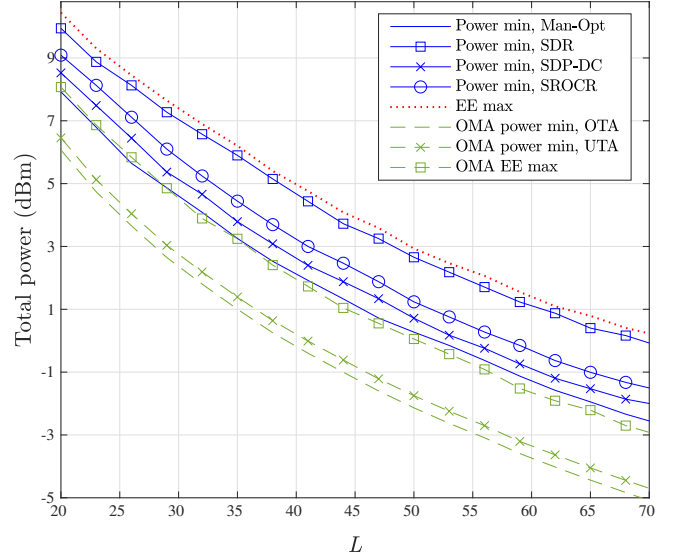


Figure 2: Sum power vs. No. of reflectors L , $R_k^{\min} = 0.2$ bits/sec/Hz

three benchmarks is much higher than the complexity of the proposed algorithm. Fig. 2 also shows that the IRS assisted OMA scheme outperforms the NOMA scheme for the case of low minimum target rate requirements for the users, hence, the IRS-OMA scheme is preferable in this case. The graph also compares the total transmit power of the users resulting from solving the sum power minimization problem against the total power resulting from solving the EE maximization problem for both NOMA and OMA cases. Clearly, and as expected, the graph shows that the sum power curve of the EE maximization problem is higher than the total power curve resulting from the power minimization problem. The rationale behind this is that the EE objective function, in (38a) for the NOMA case or in (49) for the OMA case, is pseudo-concave with respect to the transmit power vector \mathbf{p} . Hence, when the target uplink rate constraints are low, the minimum transmit powers of the users must be low too, which makes the system's EE far from the maximum EE point of the pseudo-concave objective function. Therefore, in the case of low rate constraints, moving towards the maximum EE point requires increasing the transmit powers of the users above the minimum. That is why the power consumption in the case of NOMA/OMA EE maximization is higher than that in the case of total transmit power minimization.

Fig. 3 shows the performance of the proposed EE maximization solutions using the proposed CCM based alternation maximization algorithm in Algorithm 5 discussed in Sec. IV. The IRS assisted NOMA EE maximization problem is solved by optimizing \mathbf{p} given fixed \mathbf{w} using the proposed low complexity Algorithm 4, and then optimizing \mathbf{w} given fixed \mathbf{p} alternately using the alternation maximization algorithm. When solving for \mathbf{w} given \mathbf{p} , we compare the proposed manifold optimization based feasibility expansion scheme, discussed in Sec. IV-B, against the SDR, SDP-DC and SROCR benchmarks. The figure shows that the proposed manifold optimization algorithm, when applying the feasibility expansion (FeEx.) mechanism discussed in Sec. IV-B, outperforms the three benchmarks, although it has much lower complexity. The

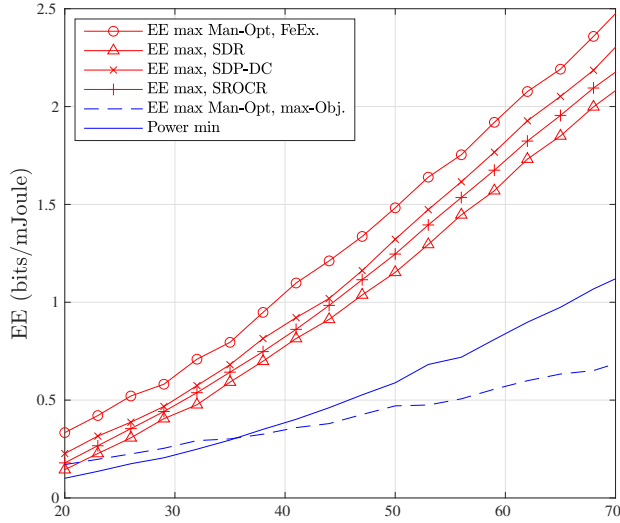


Figure 3: EE vs. No. of reflectors L , $R_k^{\min}=0.2$ bits/sec/Hz

figure also confirms that the proposed feasibility expansion mechanism provides far better performance by solving (26) than maximizing the EE cost function using (44). The proposed CCM based feasibility expansion (FeEx.) mechanism in Algorithm 5 outperforms the existing SDP based counterparts because it optimizes the original optimization vector, \mathbf{w} , on the original feasibility region as in (26). On the other hand, SDP based schemes like SDR, as in (42), broaden the original feasibility region, encompassing a larger set, \mathbf{W} , which is the outer product of the original optimization vector. When the optimal solution obtained on the larger set, in the SDP algorithms, is projected back on the original feasible region through Cholesky decomposition, the results may deviate from the optimal solution to the original problem, which causes performance reduction compared to the proposed CCM-based approach. Moreover, since the problem (37) focuses on maximizing the EE of the system, it is expected that the EE resulted from (37) is higher than the EE resulted from solving the sum power minimization problem in (8). This is because problem (8) only focuses on minimizing the transmit powers of the users without paying any attention to the EE metric, while in the EE maximization problem (37), the users are permitted to increase their transmit powers and hence, increase their uplink data rates accordingly above the minimum target rates to increase EE.

Fig. 4 compares the EE of the IRS assisted OMA scheme against the EE of the NOMA counterpart for the case of low minimum uplink rate constraints at the users. The graph shows that the resulting EE of the OMA scheme largely beats the EE of the NOMA scheme in the case of low uplink QoS requirements at the users. Fig. 4 also shows that optimizing the time allocations of the IRS assisted OMA scheme, by solving problem (49) in Sec. V-B, results in higher EE than the regular OMA scheme with equal time fractions among the users. The figure also compares the resulting EE of the OMA scheme when the objective is minimizing the total transmit power against the EE resulting from maximizing the EE of the system. Intuitively, the resulting EE of maximizing the OMA

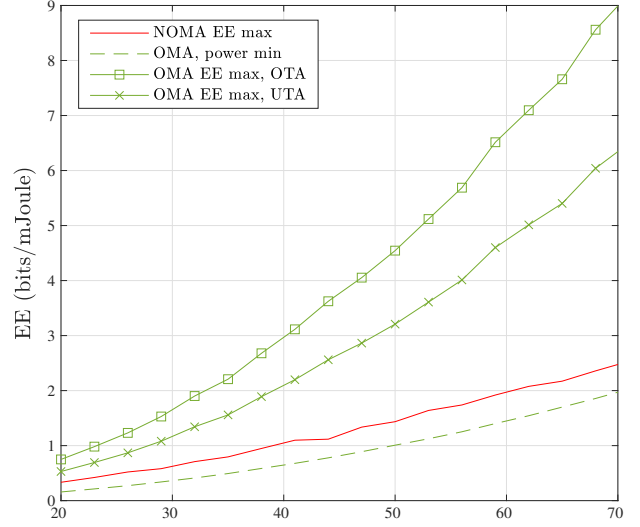


Figure 4: Comparing OMA and NOMA EE, $R_k^{\min}=0.2$ bits/sec/Hz

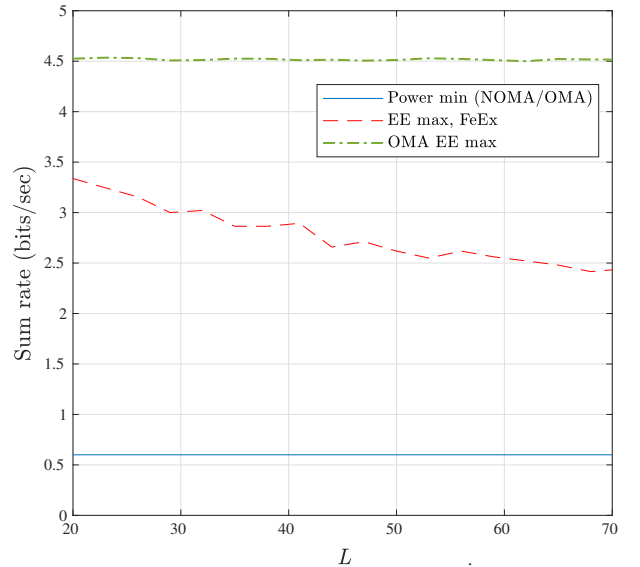


Figure 5: Sum rate vs. No. of reflectors L , $R_k^{\min}=0.2$ bits/sec/Hz

system's EE is higher than the resulting EE of minimizing the OMA total transmit power. This is because the OMA EE objective function in (49) is pseudo-concave with respect to the transmit power vector \mathbf{p} . When the target uplink rate constraints are low, the users transmit at low power levels in the case of total transmit power minimization, which makes them far from the maximum EE point. Therefore, in the case of OMA EE maximization, the users increase their transmit powers moving towards the maximum EE point of the OMA system. However, in the case of minimizing the total transmit power, we only focus on minimizing the transmit power as much as possible by satisfying the minimum target uplink rates with equality while neglecting the EE aspect.

In Fig. 5, we depict the attainable sum rates of the users within the system, attained by solving both the sum power minimization and the EE maximization problems in (8) and (37), and their OMA versions (45) and (49), respectively. When addressing the total transmit power minimization problem in a NOMA/OMA system, each user precisely meets the

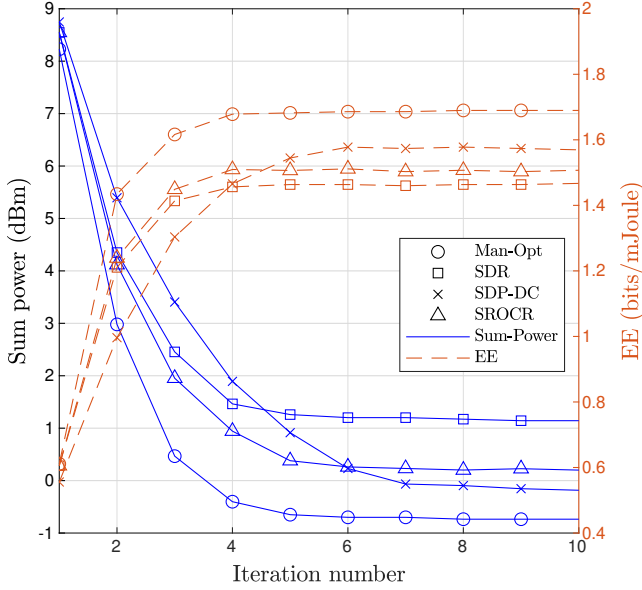


Figure 6: Convergence of sum power and EE optimization using proposed manifold and benchmarks, $R_k^{\min}=0.2$ bits/sec/Hz

required minimum uplink rate with equality. This outcome aligns with intuition, as the primary objective is to minimize the transmit powers of users as much as possible in this scenario. In contrast, when tackling the EE maximization problem in a NOMA/OMA system, users have the flexibility to transmit at higher power levels, achieving elevated uplink data rates to maximize the overall EE of the system. As illustrated in Fig. 2 and discussed in the paper, the EE cost functions for NOMA and OMA, represented by equations (38a) and (49), respectively, exhibit pseudo-concavity with respect to the transmit power vector, \mathbf{p} . Consequently, when the constraints on target uplink rates are low, users increase their transmit power, moving towards the system's maximum EE point. As a result, the achievable uplink data rates of users surpass the target minimum rates when addressing the EE maximization problem for both NOMA and OMA systems. This behavior underscores the adaptive nature of users in response to EE maximization, allowing them to dynamically adjust their transmit power levels to optimize system-wide EE.

Fig. 6 compares the convergence behaviours of the alternation optimization algorithm using the proposed manifold optimization technique against the different SDP-based benchmarks that we use in this paper. The figure shows the convergence curves of both the sum power minimization and the EE maximization problems. The graph shows that the proposed alternation optimization based manifold optimization algorithm converges after few iterations, and faster than the shown benchmarks, when solving the two considered optimization problems. These results confirm the convergence of the proposed alternation optimization algorithm in Sec. III-C.

Fig. 7 compares the time complexity of the proposed CCM algorithms in Algorithm 2 and Algorithm 5 against the SDR benchmark for solving the sum power minimization and the EE maximization problems, respectively. The figure compares both techniques for both the low and high rate constraints

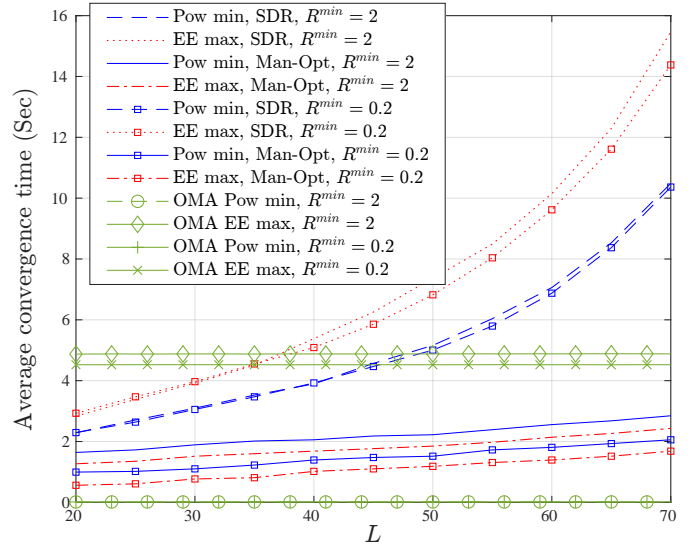


Figure 7: Time complexity of sum power and EE optimization using proposed manifold vs. SDR based techniques

cases for the sake of completeness. We plot the average run time per iteration for each optimization algorithm against the number of IRS elements. Clearly, the figure shows that the average convergence run time of the proposed CCM based optimization algorithm is far less than the average run time of the SDR benchmarks for both the sum power minimization and the EE maximization problems. These results yield a huge time complexity reduction for our proposed CCM based algorithm against the SDP-based benchmarks like SDR, which confirms the presented complexity analysis in Sec. III-D. Furthermore, we can notice from Fig. 7 that the proposed CCM algorithm has a slightly less time complexity in the low rate constraints case than the high rate case. This result makes sense since the high rate constraints make the feasible region tighter than in the low rate case, which causes the CCM to take a longer duration in searching for an optimal feasible solution. However, the difference between the low and high rate cases is nearly insignificant for the SDR benchmark. Fig. 7 also shows the time complexity of the IRS-OMA scheme discussed in Sec. V. It is noticeable from the graphs that the average convergence time of either power minimization or EE maximization of IRS-OMA does not increase with the number of IRS elements, L . This observation is anticipated since the IRS-OMA scheme does not include phase shifts optimization at IRS because they already have a closed-form solution. The complexity of the IRS-OMA power minimization is very low as shown in the figure since it is only limited by solving the low complexity problem in (47). However, the time complexity of the EE maximization of IRS-OMA is relatively high since it is determined by solving (50), which involves alternating optimization that includes applying the Dinkelbach algorithm in every step, raising the complexity of the whole algorithm.

B. NOMA vs. OMA for high minimum rate constraints

In the next figures, we show the performance of the proposed solutions to both the total transmit power minimization problem and the EE maximization problem when the minimum target uplink rates are high. We assume that the required

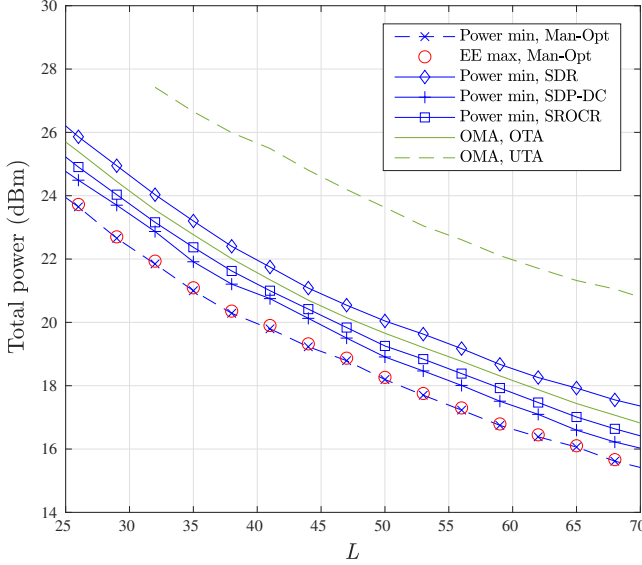


Figure 8: Sum power vs. No. of reflectors L , $R_k^{\min}=2.5$ bits/sec/Hz

minimum transmission rate for each user is 2.5 bits/sec/Hz. We will show how the two optimization problems are related when the uplink rate constraints of the users are high.

Fig. 8 presents the relationship between the total uplink transmit power of users and the number of reflecting elements for various schemes. The proposed manifold optimization scheme is compared against benchmarks such as SDR, SDP-DC, and SROCR, as well as an IRS-assisted OMA scheme. Similar to the scenario with low rate constraints, the manifold optimization-based approach in (26) outperforms the three benchmarks in the high rate constraints case too, all while maintaining significantly lower computational complexity. In contrast to the low rate constraints scenario, Fig. 8 reveals that the IRS-NOMA scheme in (8) surpasses the optimized IRS-OMA scheme in (45) when there are high minimum rate requirements for users. This implies that IRS-NOMA is more favorable than IRS-OMA when dealing with stringent minimum target rate constraints. Additionally, Fig. 8 compares the total transmit power of the users resulting from solving the sum power minimization problem against the total power from solving the EE maximization problem for both NOMA and OMA cases. Interestingly, in cases with high rate constraints, the sum power curve resulting from solving the EE maximization problem (37) perfectly coincides with the sum power curve from solving the total transmit power minimization problem (8). This behavior is explained by the pseudo-concave nature of the NOMA EE objective function in equation (38a) with respect to the transmit power vector \mathbf{p} . Consequently, when the target uplink rate constraints are high, the minimum transmit powers of the users must also be high, exceeding the maximum EE point. Therefore, in scenarios with high rate constraints, moving towards the maximum EE point requires decreasing the transmit powers of the users. Hence, maximizing EE directly leads to minimizing the total transmit power of users, resulting in the overlapping of both solutions.

Fig. 9 shows a performance comparison between the proposed EE maximization solutions using the alternating max-

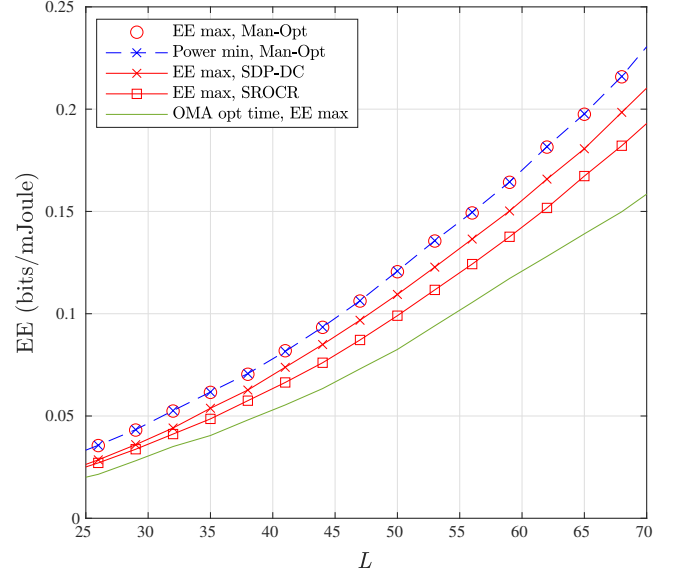


Figure 9: EE vs. No. of reflectors L , $R_k^{\min}=2.5$ bits/sec/Hz

imization scheme discussed in Sec. IV for the case of high uplink rate requirements at the users. Similar to the case of low rate constraints at the users, the figure shows the performance gain of the obtained EE using the proposed manifold optimization over the SDR, SDP-DC and SROCR benchmarks. It is clear from Fig. 9 that the EE resulting from minimizing the total transmit power by solving (8) is exactly the same as the EE resulting from maximizing the EE of the system by solving (37). This supports the results and discussion of Fig. 8 since we found that the EE maximization problem coincides with the total transmit power minimization. Therefore, the two problems lead to the same EE values in the case of high uplink rate requirements.

VII. CONCLUSIONS AND FUTURE WORK

In this article, we proposed efficient solutions to the sum power minimization and the EE maximization problems for IRS-assisted uplink NOMA networks. The two problems are solved by jointly optimizing the users' uplink transmit powers and the IRS passive beamforming coefficients using alternating optimization-based algorithms. During the passive beamforming sub-problem, we showed that the proposed manifold optimization solutions have significant performance gains and are far superior to the SDR benchmark solutions. Additionally, we compared our solutions for the IRS assisted NOMA system with the IRS-assisted OMA counterpart.

The simulation results indicate that NOMA exhibits superior performance over OMA when the users' uplink data rate requirements are high. Conversely, OMA is more favorable when the rate requirements are low. This behavior can be attributed to the fact that there is an optimal closed-form solution for the phase shifts of IRS elements with OMA signaling, which provides some performance gain over IRS-NOMA. This gain might overwhelm the gain obtained by IRS-NOMA when the minimum data rate requirements are low because the spectral efficiency gained by NOMA signaling is

negligible under these operating conditions as the achievable rate of the near user is much higher than the other users. On the other hand, when the minimum rate requirements are high, the achievable rates for all users will be comparable, which implies that the system operates in the mid-capacity region in which NOMA signaling has a superior gain. This conclusion does not actually undermine the benefits of IRS-NOMA since it can be used in applications that require relatively high data rates such as UAV communications and IoT data collection points. On the other hand, IRS-OMA is useful for low-rate applications such as wireless sensor networks (WSN).

In future work, it is worth investigating the performance of the developed low-complexity CCM based algorithm to optimize uplink NOMA systems with multiple antennas at the BS. It will be interesting to alternatively optimize the IRS phase shifts, using the CCM algorithm, the transmit powers of the users, and the NOMA receive beamforming vector.

REFERENCES

- [1] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Trans. Cognitive Commun. Netw.*, vol. 6, no. 3, pp. 990–1002, 2020.
- [2] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.
- [3] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, J. de Rosny, A. Bounceur, G. Lerosey, and M. Fink, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, 2019.
- [4] Y.-C. Liang, R. Long, Q. Zhang, J. Chen, H. V. Cheng, and H. Guo, "Large intelligent surface/antennas (LISA): Making reflective radios smart," *J. Commun. Inf. Netw.*, vol. 4, no. 2, pp. 40–50, 2019.
- [5] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, 2021.
- [6] 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, 2019.
- [7] 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, 2019.
- [8] D. Zhao, H. Lu, Y. Wang, H. Sun, and Y. Gui, "Joint power allocation and user association optimization for 5G-aided mmwave systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 577–590, 2022.
- [9] M. Al-Jarrah, E. Alsusa, A. Al-Dweik, and D. K. C. So, "Capacity analysis of IRS-based UAV communications with imperfect phase compensation," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1479–1483, 2021.
- [10] X. Zhou, S. Yan, Q. Wu, F. Shu, and D. W. K. Ng, "Intelligent reflecting surface (IRS)-aided covert wireless communications with delay constraint," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 532–547, 2022.
- [11] M. Alaaeldin, E. A. Alsusa, and K. G. Seddik, "IRS-assisted physical layer network coding over two-way relay fading channels," *IEEE Trans. Veh. Technol.*, pp. 1–1, 2022.
- [12] M. A. Al-Jarrah, E. Alsusa, A. Al-Dweik, and M.-S. Alouini, "Performance analysis of wireless mesh backhauling using intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3597–3610, 2021.
- [13] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, 2017.
- [14] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surv. Tutor.*, vol. 20, no. 3, pp. 2294–2323, 2018.
- [15] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017.
- [16] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann, "Modulation and multiple access for 5G networks," *IEEE Commun. Surv. Tutor.*, vol. 20, no. 1, pp. 629–646, 2018.
- [17] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, 2017.
- [18] Y. Liu, H. Xing, C. Pan, A. Nallanathan, M. Elkashlan, and L. Hanzo, "Multiple-antenna-assisted non-orthogonal multiple access," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 17–23, 2018.
- [19] F. Wang, J. Xu, and Z. Ding, "Multi-antenna NOMA for computation offloading in multiuser mobile edge computing systems," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2450–2463, 2019.
- [20] R. Duan, J. Wang, C. Jiang, H. Yao, Y. Ren, and Y. Qian, "Resource allocation for multi-UAV aided IoT NOMA uplink transmission systems," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7025–7037, 2019.
- [21] X. Mu, Y. Liu, L. Guo, J. Lin, and Z. Ding, "Energy-constrained UAV data collection systems: NOMA and OMA," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6898–6912, 2021.
- [22] M. Zeng, X. Li, G. Li, W. Hao, and O. A. Dobre, "Sum rate maximization for IRS-assisted uplink NOMA," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 234–238, 2021.
- [23] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communication: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, 2020.
- [24] S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M.-S. Alouini, "Deep denoising neural network assisted compressive channel estimation for mmwave intelligent reflecting surfaces," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9223–9228, 2020.
- [25] A. Taha, Y. Zhang, F. B. Mismar, and A. Alkhateeb, "Deep reinforcement learning for intelligent reflecting surfaces: Towards standalone operation," in *2020 IEEE 21st Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, 2020, pp. 1–5.
- [26] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 898–912, 2022.
- [27] Y. Wang, H. Lu, and H. Sun, "Channel estimation in 5G-enhanced mmwave system with super-resolution network," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2599–2603, 2021.
- [28] Y. Feng, J. Chen, X. Xue, K. Wu, Y. Zhou, and L. Yang, "Max-min fair beamforming for IRS-aided secure NOMA systems," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 234–238, 2022.
- [29] C.-W. Chen, W.-C. Tsai, and A.-Y. Wu, "Low-complexity two-step optimization in active-IRS-assisted uplink NOMA communication," *IEEE Commun. Lett.*, pp. 1–1, 2022.
- [30] L. Cantos, M. Awais, and Y. H. Kim, "Max-min rate optimization for uplink IRS-NOMA with receive beamforming," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2022.
- [31] G. Li, M. Zeng, D. Mishra, L. Hao, Z. Ma, and O. A. Dobre, "Energy-efficient design for IRS-empowered uplink MIMO-NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9490–9500, 2022.
- [32] G. Chen, Q. Wu, C. He, W. Chen, J. Tang, and S. Jin, "Active IRS aided multiple access for energy-constrained IoT systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1677–1694, 2023.
- [33] J. Na, J. Kang, and J. Kang, "Intelligent reflecting surface-assisted uplink NOMA for eMBB and URLLC coexistence," *IEEE Trans. Veh. Technol.*, pp. 1–6, 2023.
- [34] J. Li and Y. Hong, "Design of intelligent reflecting surface phase shifts for uplink SCMA," *IEEE Commun. Lett.*, vol. 27, no. 5, pp. 1437–1441, 2023.
- [35] G. Yang, X. Xu, and Y.-C. Liang, "Intelligent reflecting surface assisted non-orthogonal multiple access," in *2020 IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2020, pp. 1–6.
- [36] 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, 2020.
- [37] F. Fang, Y. Xu, Q.-V. Pham, and Z. Ding, "Energy-efficient design of IRS-NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 14 088–14 092, 2020.
- [38] M. Fu, Y. Zhou, Y. Shi, and K. B. Letaief, "Reconfigurable intelligent surface empowered downlink non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3802–3817, 2021.

- [39] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, 2020.
- [40] X. Yu, D. Xu, and R. Schober, "MISO wireless communication systems via intelligent reflecting surfaces : (invited paper)," in *2019 IEEE/CIC Int. Conf. Commun. in China (ICCC)*, 2019, pp. 735–740.
- [41] M. A. ElMossallamy, K. G. Seddik, W. Chen, L. Wang, G. Y. Li, and Z. Han, "RIS optimization on the complex circle manifold for interference mitigation in interference channels," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6184–6189, 2021.
- [42] A. Agarwal, R. Chaurasiya, S. Rai, and A. K. Jagannatham, "Outage probability analysis for NOMA downlink and uplink communication systems with generalized fading channels," *IEEE Access*, vol. 8, pp. 220461–220481, 2020.
- [43] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.
- [44] M. Zeng, N.-P. Nguyen, O. A. Dobre, Z. Ding, and H. V. Poor, "Spectral- and energy-efficient resource allocation for multi-carrier uplink NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 9293–9296, 2019.
- [45] Q. Wu, X. Zhou, and R. Schober, "Irs-assisted wireless powered noma: Do we really need different phase shifts in DL and UL?" *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1493–1497, 2021.
- [46] Y. Cheng, K. H. Li, Y. Liu, K. C. Teh, and H. Vincent Poor, "Downlink and uplink intelligent reflecting surface aided networks: NOMA and OMA," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3988–4000, 2021.
- [47] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. Springer Publishing Company, Incorporated, 2015.
- [48] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [49] Z.-q. Luo, W.-k. Ma, A. M.-c. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, 2010.
- [50] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press, 2008.
- [51] M. C. Pinar and S. A. Zenios, "On smoothing exact penalty functions for convex constrained optimization," *SIAM J. Optimization*, vol. 4, no. 3, 1994.
- [52] C. Liu and N. Boumal, "Simple algorithms for optimization on riemannian manifolds with constraints," *Appl Math Optim.*, vol. 82, no. 3, pp. 949–981, 2020.
- [53] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a MATLAB toolbox for optimization on manifolds," *J. Machine Learning Research*, vol. 15, no. 42, pp. 1455–1459, 2014. [Online]. Available: <https://www.manopt.org>
- [54] Z.-q. Luo, W.-k. Ma, A. M.-c. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, 2010.
- [55] 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, 2018.



Mahmoud AlaaEldin (Member, IEEE) received the B.Sc. degree in electrical engineering from Alexandria University, Alexandria, Egypt, in 2013, and the M.Sc. degree in electronics and communications engineering from the American University in Cairo (AUC), Cairo, Egypt, in 2019. He received his Ph.D. degree in electrical and electronics engineering from The University of Manchester, Manchester, UK, in 2024. Dr. AlaaEldin received a University Fellowship from AUC for his M.Sc., and the Marie Skłodowska-Curie Actions (MSCA) fellowship from the European Union (EU), under the EU's Horizon 2020 research and innovation program, for his Ph.D. His research interests span diverse topics in wireless communications and signal processing including multi-user MIMO/OFDM systems, massive MIMO systems, channel estimation and feedback, intelligent reflecting surfaces (IRSs), integrated radar and communications, near field communications, optimization, physical layer network coding, and applications of machine learning in wireless communications.



Emad Alsusa (Senior Member, IEEE) is a Fellow of the UK Higher Academy of Education. He earned his PhD in Telecommunications from the University of Bath, UK, in 2000. That same year, he began working on high data rate systems within an industrial project at Edinburgh University. He joined Manchester University in September 2003 where his current position is a Professor of Wireless Communications and Signal Processing and the Head of the Communications Research Group. Emad has authored over 300 journals and refereed conference publications, prominently featured in top IEEE transactions and conferences. He serves as an Editor for IEEE Wireless Communications Letters, the IEEE Transactions on Vehicular Technology. He has held roles as TPC Track Chair for conferences such as VTC'16, GSN'16, PIMRC'17, and Globecom'18, and served as General Co-Chair for the OnlineGreenCom'16 Conference. Emad has received several awards, including the Best Paper Award at the International Symposium on Power Line Communications 2016 and the Wireless Communications and Networks Conference 2019.



Karim G. Seddik (Senior Member, IEEE) received the B.Sc. (Hons.) and M.Sc. degrees in electrical engineering from Alexandria University, Alexandria, Egypt, in 2001 and 2004, respectively, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2008. He is currently a Professor in the Electronics and Communications Engineering Department at the American University in Cairo (AUC), and an Associate Dean of graduate studies and research with the School of Sciences and Engineering (SSE), AUC. He was a recipient of the American University in Cairo Faculty Merit Award for Excellence in Research and Creative Endeavors in 2021. He was a recipient of the State Encouragement Award in 2016 and the State Medal of Excellence in 2017. He also co-authored a conference paper that received the Best Conference Paper Award from the IEEE Communication Society Technical Committee on Green Communications and Computing in 2019.



Mohammad Al-Jarrah (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering/wireless communications from Jordan University of Science and Technology (JUST), Irbid, Jordan, in 2008 and 2011, respectively. He received his Ph.D. degree in electrical and electronics engineering from The University of Manchester, Manchester, U.K in 2023. Dr. Al-Jarrah is a recipient of the Royal Jordanian Scholarship and Marie Skłodowska-Curie Actions (MSCA) fellowship from the European Union (EU). His research interests include diverse topics in wireless communications and signal processing: statistical signal processing, localization and tracking, RFID communications, spectrum sensing, integrated radar-communications, intelligent reflecting surfaces (IRSs), and backhauling and cellular planning for future communication networks.



Dr. Constantinos B. Papadias (Fellow, IEEE) is the Associate Vice President for Research, Innovation and Graduate Studies of The American College of Greece (ACG) in Athens, Greece. At ACG, he is also the Executive Director of the Research, Technology and Innovation Network (RTIN), Head of the Smart Wireless Future Technologies (SWIFT) Lab and Professor of Information Technology. Prior to these, he was the Dean of Athens Information Technology (AIT), in Athens, Greece, where he was also Professor and Head of the Broadband Wireless and Sensor Networks (B-WiSE) Research Group. He has held Adjunct Professorships at several Universities, currently with Aalborg University and with the University of Cyprus. He has published over 240 scientific papers, 12 patents and 4 books and has received more than 10000 citations and several recognitions for his work, including 4 best paper awards. He has participated in and led several EU-funded research projects, including his current role as Scientific Manager in the Doctoral Network project ISLANDS, and his recent roles as Research Coordinator in the ITN Project PAINLESS (2018-2022) and Technical Coordinator in the EU CHIST-ERA project FIREMAN. He currently serves as Director-at-Large of the IEEE Signal Processing Society for Region 8 (Europe, Middle East and Africa).