


Joint Beamforming and Aerial IRS Positioning Design for IRS-Assisted MISO System With Multiple Access Points

TANG CHAO¹, CARRSON C. FUNG¹ ² (Member, IEEE), ZI-EN NI² (Student Member, IEEE), AND MYKOLA SERVETNYK² (Member, IEEE)

¹Department of Communications and Information Engineering, School of Engineering, Tokyo Institute of Technology, Tokyo 152-8550, Japan

²Institute of Electronics, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan

CORRESPONDING AUTHOR: C. C. FUNG (e-mail: c.fung@ieee.org)

This work was supported in part by the Google exploreCSR Project under Grant 112Q90004C.

ABSTRACT Intelligent reflecting surface (IRS) is a promising concept for 6G wireless communications that allows tuning of the wireless environments to increase spectral and energy efficiency. Many optimization techniques have been proposed in literature to deal with the joint passive and active beamforming design problem, but without any optimality guarantees for the multiple access points (APs), multiple IRSs, and multiple users scenario. Moreover, the multiple access problem is also considered with the beamformer design which has not been addressed in literature, except in the context of joint transmission, which is not considered herein. To further maximize ground based and support non-terrestrial communications, the joint aerial IRS (AIRS) positioning and beamformer design problem is also considered. In the first part of the paper, an algorithm considering predefined AP-user pairing is proposed, which allows beamforming vectors to be designed distributively at each access point by using Generalized Bender Decomposition (GBD), consequently resulting in certain level of optimality. The problem can be transformed via mathematical manipulation and semidefinite relaxation (SDR) into a convex problem and solve using semidefinite programming (SDP). Another algorithm was developed to solve for optimal AP-user pairing at the same time by introducing additional binary variables, making the problem into a mixed-integer SDP (MISDP) problem, which is solved using GBD-MISDP solver, albeit with higher computational and time complexity than the GBD for the original problem. A heuristic pairing algorithm, called GBD-iterative link removal (GBD-ILR), is proposed to combat this problem and it is shown to achieve solution close to that of the GBD-MISDP method. A joint AIRS positioning and beamformer design problem is solved in the second part by using the proposed successive convex approximation-alternating direction of method of multipliers-GBD (SAG) method. Simulation results show the effectiveness of all proposed algorithms for joint beamformer design, joint beamformer design with AP-user pairing in a multiple access points system, and the joint AIRS positioning and beamformer design. In addition to simulation results, an analysis of communication overhead incurred due to use of the IRS is also given.

INDEX TERMS Intelligent reflecting surface (IRS), aerial IRS (AIRS), beamforming design, generalized benders decomposition, mixed integer programming, semidefinite relaxation, distributed wireless system design.

I. INTRODUCTION

INTELLIGENT reflecting surface, or IRS, has recently been touted as a low cost, but effective, technology for 6G wireless communications due to its ability to direct signal

to different directions and manipulate channel condition without excessive power consumption [1], [2], [3], [4]. This is a key enabler for new bands (e.g., mmWave) as increase signal propagation loss severely limits coverage

area. In an IRS-assisted system, it is envisioned that the transmitter will control the IRSs that are mounted on either facade of buildings, UAVs, or indoor on the ceiling, making transmission possible despite existence of blockage between the transmitter and receiver. An IRS comprises of an array of antenna elements, each of which can independently invoke some change to the incident signal. In most studies, the change of the incident signal is considered a phase shift and amplitude reduction only.

In addition to the aforementioned advantages, recent studies have proposed various ways of integrating IRS in the current infrastructure. Given that IRS can be used with full-duplex transmission without self-interference, it is a potential replacement for the current state-of-art solutions, such as amplify-and-forward relay and backscatter communication techniques. Moreover, IRS can be effectively used for interference cancellation. It has been shown in [5] that IRS can significantly improve received signal power with only slightly interference growth in a multi-cell system, resulting in higher SINR. Unlike active elements such as relays, passive IRS also scales well with higher frequencies. In addition, as passive IRS simply reflects signals, it can be easily incorporated with current solution to create active-passive hybrid networks, and lower both the energy consumption and cost of the current solutions [5], [6].

Some approaches have been proposed in literature for the joint beamformer design problem at the APs and IRSs. In [7], an algorithm is proposed for single AP, single user equipment (UE) MISO system aided by a single IRS panel to maximize the received signal power. A formulation for single AP multi-user MISO systems is also proposed, where the goal is to minimize the transmit power while constraining the signal-to-interference plus noise ratio (SINR) above certain level to ensure quality of service. Both problems are solved using semidefinite relaxation and alternating optimization. Maximum ratio transmission solution was proposed for the single user case while a minimum mean-square error precoding strategy was used in the multi-user case. Many other works similarly considered the single AP, multi-user MISO system aided by a single IRS panel [8], [9], [10], [11], [12], [13]. Ma et al. in [8] aims to maximize the sum rate and provides a low complexity solution. The authors first decomposed the original formulation into three disjoint subproblems by applying fractional programming (FP) techniques [14], [15], and derive closed-form solutions for each of the subproblems. Both perfect and imperfect CSI scenarios were considered. In [9], Guo et al. maximized the weighted sum rate of the system by decomposing the original problem into two subproblems via Lagrangian dual transform. The precoders and phase shifters are then obtained by optimizing the two subproblems alternatively. A closed-form solution for the phase shifters was also proposed to lower the complexity. In [10], by applying both FP techniques and the alternating direction method of multipliers (ADMM) algorithm, Di et al. proposed an iterative solution to maximize the sum rate of the system

while considering the practical case of discrete phase shifters at the IRS. In [11], Chen et al. maximized the minimum secrecy rate to achieve physical layer security. By applying alternating optimization and path-following algorithm, an iterative solution is obtained. The authors were able to further lower the computational complexity of the algorithm by assuming zero-forcing (ZF) precoding. Huang et al. in [12], [13] aims to maximize the energy efficiency. By employing ZF precoding with power allocation at the APs, the precoder design problem reduces to the power allocation problem at the AP and becomes decoupled with the phase shifter design problem. Alternating optimization is then applied. However, ZF precoding is known to introduce noise enhancement, as well as a severely compromising performance over ill-conditioned channels. The algorithm also cannot be applied to other precoding schemes.

For other system setups, Li et al. in [16] considered single AP, multiple UEs MISO systems with multiple IRS panels and proposed an iterative solution to maximize weighted sum rates. The authors first converted the original problem into a weighted minimum mean-square error (WMMSE) problem, then applied alternating optimization and Riemannian manifold conjugate gradient (RMCG) method. Unfortunately, no proof of convergence of this algorithm is provided. Considering single IRS-assisted MIMO systems with single AP and single UE, Wang et al. in [17] proposed an iterative solution to maximize ergodic achievable rates. First, a tight upper bound was derived and alternating optimization was used to derive two subproblems. By assuming certain spatial channel model, the authors were able to reformulate both problems into SDPs. Instead of relying on current channel state, the solution requires knowledge of the statistical property of the channel, which may be difficult to obtain in practice. Also, the convergence of the proposed algorithm is not guaranteed.

Aside from optimization techniques, deep learning-based approaches have also been utilized to solve the joint beamforming problem. Taha et al. in [18] proposed using a hybrid IRS that is capable of estimating channels to achieve efficient phase shifter designs. Two solutions using the observed sampled channel were proposed. The first one estimated the complete channel with compressive sensing techniques and then looked up the phase shifter solutions in a codebook. The second directly mapped the sampled channel to the phase shifter solutions with a multi-layer perceptron model, which was trained with the goal to maximize the achievable rate in a single AP, single UE MISO system. Lin et al. in [19] considered the joint beamforming problem for single AP, single UE MISO systems under uncertain channel conditions and formulated a transmit power minimization problem with the quality of service constraints. To solve the problem, a deep reinforcement learning (DRL) based approach called deep deterministic policy gradient (DDPG) algorithm was applied. Specifically, a convex approximation was derived as a lower bound of the original problem and then used in the DRL framework. The passive beamformers are

obtained through the DDPG algorithm, while the precoders are obtained subsequently using the model-based convex approximation. In [20], Huang et al. similarly adopted the DRL framework to obtain the precoders and the phase shifters simultaneously for single AP multiple UE MISO systems. Albeit the promising results, deep learning-based solutions usually require extensive training time and samples to obtain a solution and thus can be time consuming and computationally expensive. Moreover, it is often difficult to quantify their optimality as the design approaches are usually data-driven.

For joint transmission coordinated multipoint systems, Huang et al. in [21] considered a multiple AP multiuser MISO system assisted by multiple IRSs with joint transmission. The authors solved the resulting rate maximization problem by reformulating the fractional programming problem into a form that follows the ADMM algorithm, which leads to a complete decentralized solution where the precoders at the APs are designed distributively and locally. Alternatively, Zhang et al. in [23] formulated the joint beamformer design problem as an energy efficiency maximization problem for a more general system with multiple IRSs and APs, and single antenna UEs. The formulated problem is simplified through adopting ZF beamforming with power allocation at the APs. Alternating optimization is then applied to create an iterative algorithm. Since joint transmission is used, the interference are eliminated through the ZF beamformer, which has the aforementioned drawbacks.

Both [7] and [16] have shown that solving beamforming vectors and IRS reflection coefficients jointly, while creating a nonconvex problem, is beneficial to overall system performance. However, approaches taken in previous works, rely on alternating optimization which usually does not converge and often suboptimal. In this work, a distributed algorithm using Generalized Benders decomposition (GBD), which deals with coupling between IRS and AP, is proposed. The problem is formulated by assuming underlay spectrum sharing is used in which the instantaneous leakage interference power of each pair of transceivers is limited by a certain leakage threshold that is computed based on the standardization document [24]. This complements well with non-orthogonal multiple access techniques to boost massive connectivity in future wireless systems [22]. Two resulted subproblems are further solved efficiently via the interior-point method applied to its convex reformulations. Next, the predefined AP-UE pairing assumed previously is relaxed, which requires solving the pairing problem along with the joint beamforming problem. The problem can be formulated as a mixed-integer semidefinite programming problem (MISDP) by introducing a binary variable indicating AP-UE pairing that can be solved using a GBD-MISDP solver, albeit incurring high computational and time complexity. A heuristics algorithm is proposed and shown to achieve solution close to that of the GBD-MISDP method with lower computational and time complexity. Besides simulation results to establish the efficacy of the proposed methods, an

analysis on the communication overhead due to the use of IRSs is also given.

Due to the need to further optimize communication performance for ground based networks or non-terrestrial networks, a joint aerial IRS (AIRS) and beamformer design problem is also considered in this work. Wu and Zhang in [7] discussed about the positioning of the AIRS and how it affects the performance of the system, but the authors never proposed an algorithm to determine the optimal position of the IRS. Zeng et al. in [25] considered a single AP and single UE MISO AIRS-assisted system that optimized the cell coverage by optimizing the horizontal distance between the AP and AIRS, and the orientation of the AIRS. It was shown that when the AIRS faces toward the AP vertically, the service coverage can be maximized. Several other works [26], [27], [28], [29], [30], [31], [32] considered the scenario where the IRS is mounted on a unmanned aerial vehicle (UAV). Reference [26] considered the joint AIRS and UE positioning and joint beamformer design problem for a MISO AIRS system with single AP and single UE. Since only a single UE is involved, the active beamformer simply equals to the maximum ratio transmission (MRT) precoder, where the authors tried to maximize the worst-case SNR. The solution for the passive beamformer can be obtained by first fixing the location of the AIRS and the solution guarantees that all reflected signals are combined coherently at the location of the UE. Unfortunately, the approach does not extend to the multiple APs and UEs scenario. [27] considered an AIRS-assisted MISO system with multiple APs jointly serving a single UE. The solution for the position of the AIRS is obtained by maximizing the sum rate. The problem is simplified by neglecting all APs except for the one closest to the UE in order to obtain a closed-form solution. Reference [28] solved the joint AIRS trajectory and beamformer design problem using successive convex approximation (SCA). A similar problem is solved in [29] with the added complexity of finding a solution for the user scheduling problem. The authors devised an iterative alternating approach which divided the entire problem into several subproblems, but no convergence analysis was given. Another iterative alternating approach to solve the user scheduling, AIRS trajectory and beamformer problem was proposed in [30], but the power consumption is minimized instead. Even though the algorithm is guaranteed to produce a suboptimal solution, the system considered only contains a single AP. In [31], the optimal horizontal position of the AIRS is first determined by minimizing the path loss between the AP and AIRS, and between the AIRS and UEs in a MISO AIRS system, followed by solving for the optimal passive and active beamformer solutions that enhance the performance of the strong user while maintaining a certain level of quality of service for the weaker user. However, the authors only considered a two-user system, and the approach will lead to a suboptimal solution for both the position and beamformers as they are not solved iteratively. Finally, [32] aimed to minimize the number of antenna elements on the

AIRS in a MISO AIRS system with a single AP and UE while determining the optimal passive beamformer.

In this work, the SAG method is proposed to jointly design the AIRS position, and the passive and active beamformers for MISO AIRS-assisted systems with multiple APs and single-antenna UEs. The problem is divided into different subproblems and solved iteratively. Even though one of the steps involved solving a nonconvex problem (using gradient ascent plus momentum method) [33] so that convergence is not guaranteed, the method always converges with a reasonable initial point, i.e., a point near the APs as it has been shown in different literature [26], [27] that the optimal solution is often near the APs.

The main contributions of this work are as follows:

- This work jointly optimized phase shifters at multiple IRSs and linear precoders at the APs by solving a nonconvex programming problem under the underlay spectrum sharing scenario. The resulting solution directly maximizes received signal power at the UEs, which in return increases the overall system sum rate. This can be viewed as expanding the works in [7] and [16] to multi-AP scenarios.
- Proposed GDB-based algorithm features nice convergence properties and separability of subproblem, which is important for the practical implementation. Moreover, resulting solutions are guaranteed to be nearly optimal, i.e., converge to the ε -optimal solution within finite number of iterations compared to asymptotically optimal solutions in other methods.
- The problem of AP-UE pairing in IRS-assisted system is considered and incorporated into the original joint beamforming problem. The problem is formulated and solved using the proposed GBP-MISDP algorithm such that the pairing solution is guaranteed to be the global optimal, albeit with increase in computational and time complexity. A heuristic approach, called GBD-iterative link removal (GBD-ILR), is proposed to deal with these problems and is shown in internal experiments to achieve received signal power performance close to that of the GBP-MISDP method. The sum rate performance of the GBD-ILR vs. GBD-MISDP is shown in the sequel.
- The advantages of the proposed GBD algorithm over alternating maximization and fixed IRS phase shifters is shown. The proposed algorithm is also compared against the solution in [16] in a single IRS setup. Comparison with [7] is difficult, and therefore left out, as it was difficult to determine the QoS threshold to have a fair comparison. Simulation results using the GBD-MISDP and the proposed heuristic algorithm for solving the AP-UE pairing problem are also given and shown that the latter algorithms can obtain solutions close to that of the GBD-MISDP. Finally, analysis on the amount of communication overhead for the IRS-assisted system is also provided.

- A joint AIRS positioning, passive and active beamformer design algorithm, called SAG, is also proposed, for the multi-AP, multiple single-antenna UEs AIRS-assisted system that aims to maximize the received signal power while capping the leakage interference. Even though there is no guarantee of convergence, using the gradient ascent with momentum method always attain successful convergence and a solution that is better than the benchmarks in all of the trials. Unlike other methods tested, the proposed SAG method can render optimal 3D location of the AIRS. However only 1D and 2D results are provided.
- Simulation of the GBD-based passive and active beamformer with AP-UE pairing was done under the indoor environment and the channel model stated in the standard documentation [34]. Simulation of the SAG algorithm for the joint AIRS positioning, and passive and active beamformer design is done in an outdoor environment and channel model in [34].
- Comparison with the exhaustive 1D and 2D search methods, and the method in [27], is carried out with the proposed SAG method. Since both benchmarks did not include precoder design, the zero-forcing precoder and the proposed GBD-based precoder design were used in conjunction with both methods.

The rest of the paper is organized as follows. A description of the system model and problem formulation is given in Section II. The proposed GBD based algorithm for solving the joint beamforming problem is described in Section III. The GBD-MISDP problem and solution are provided in Section IV. This is followed by a complete description of the proposed heuristic AP-UE pairing algorithm in Section V called GBD-ILR (GBD-iterative link removal). The proposed SAG method is described Section VI. Simulation results with communication overhead analysis are provided in Section VII. The paper is concluded in Section VIII.

Notations: Uppercase (lowercase) bold face letters indicate matrices (column vectors). Superscript H denotes Hermitian, T denotes transposition. $\mathbf{A} \succeq 0$ designates \mathbf{A} as a symmetric positive semidefinite matrix, $\mathbf{1}_M$ denotes an $M \times 1$ vector, containing 1 in all of its entries. $[\mathbf{A}]_{ij}$ denotes the (i, j) th element of \mathbf{A} ($[\mathbf{a}]_i$ is defined similarly for vector \mathbf{a}). $\|\cdot\|_2$ and $\|\cdot\|_F$ denote ℓ_2 and Frobenius norms, respectively. $|\mathbf{A}|$ denotes the elementwise magnitude value of \mathbf{A} . $\text{Diag}(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ is diagonalization operator, i.e., its output is a square matrix with operator's argument being the main diagonal. $\text{diag}(\cdot) : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^n$ denotes an inverse operation to $\text{Diag}(\cdot)$. $10^{\mathbf{D}}$ denotes a diagonal matrix with the i th diagonal element equals to $10^{[\mathbf{D}]_{ii}}$, where \mathbf{D} is a diagonal matrix. Operator $\lceil \cdot \rceil$ denotes ceiling operator which rounds the argument to the nearest greatest integer. \mathbf{x}^* denotes the optimal solution of a problem.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a system consisting of Q APs, each with N_T transmit antennas, I single-antenna UEs, and L IRSs each

with M reflecting elements. $\mathcal{Q} = \{1, \dots, Q\}$, $\mathcal{I} = \{1, \dots, I\}$, and $\mathcal{L} = \{1, \dots, L\}$ denote the set of APs, UEs, and IRSs, respectively. It is assumed at first that the UEs have already been paired with the APs, and that each AP will serve approximately the same number of users. This is known as random pairing (RP). This will be relaxed in Sections IV and V. Let \mathcal{P} be a set that contains the AP-UE pairs (q, i) , indicating the i th UE is served by the q th AP. Each UE is only served by a single AP and all APs are assumed to be active.

In this work, subscript a,b is used to denote “from a to b ”. For instance, $\mathbf{H}_{q,\ell}$ denotes the channel from the q th AP to the ℓ th IRS, and $\mathbf{h}_{\ell,i}$ denotes the channel from the ℓ th IRS to the i th UE. $\mathbf{f}_{q,i}$ denotes the active beamforming vector designed for signal transmission from the q th AP to i th UE, through one of the IRSs. Define $\boldsymbol{\phi} \triangleq [\phi_1^1 \cdots \phi_M^1 \phi_1^2 \cdots \phi_M^L]^T$ and $\boldsymbol{\beta} \triangleq [\beta_1^1 \cdots \beta_M^1 \beta_1^2 \cdots \beta_M^L]^T$, with $\beta_m^\ell \in [0, 1]$, as \mathbb{C}^{ML} vectors, containing phase shifts and magnitude suppression factors at each IRS elements, respectively. Further define $\boldsymbol{\Theta} \triangleq \text{Diag}([\beta_1^1 \exp(j\phi_1^1) \cdots \beta_M^1 \exp(j\phi_M^1) \beta_1^2 \exp(j\phi_1^2) \cdots \beta_M^L \exp(j\phi_M^L)]) \in \mathbb{C}^{ML \times ML}$ as a diagonal matrix which mathematically models the effect the IRS has on the incident signal. $\mathbf{H}_{q,R} \triangleq [\mathbf{H}_{q,1}^T \cdots \mathbf{H}_{q,L}^T]^T \in \mathbb{C}^{ML \times N_T}$ and $\mathbf{h}_{R,i} \triangleq [\mathbf{h}_{1,i} \cdots \mathbf{h}_{L,i}] \in \mathbb{C}^{1 \times ML}$ denote the aggregate channel matrix and vector from the q th AP to the IRSs (denoted as R) and from the IRSs to the i th user, respectively. $\mathbf{H}_{q,\ell} \in \mathbb{C}^{M \times N_T}$ and $\mathbf{h}_{\ell,i} \in \mathbb{C}^{1 \times M}$, for $\ell = 1, \dots, L$, denote the channel matrix and vector as defined above. It shall be assumed hereafter that 1) the direct link between the APs and UEs are blocked, implying that the APs can only reach their respective users using the IRSs, and 2) the APs reach the UEs using only a single-hop reflection off of one of the IRSs, i.e., no signals are reflected off of one IRS to another. The received signal at the i th UE can then be expressed as $y_i = \mathbf{h}_{R,i} \boldsymbol{\Theta} \mathbf{H}_{q,R} \mathbf{f}_{q,i} s_{q,i} + \sum_{\{(q,j)|(q,j) \in \mathcal{P}, j \neq i\}} \mathbf{h}_{R,i} \boldsymbol{\Theta} \mathbf{H}_{q,R} \mathbf{f}_{q,j} s_{q,j} + n_i$, where $s_{q,i}$ is the signal from the q th AP to the i th UE with $\mathbb{E}[|s_{q,i}|^2] = 1$ and n_i is the AWGN noise with variance σ^2 . The first term is the received signal, the second term is the interference, and the third term is the AWGN noise. Due to the monotonic relationship between received signal power and SINR, the direct dependence of achievable throughput on the SINR, and to make finding the solution easier, the design problem considers maximizing the received signal power while constraining the pairwise instantaneous leakage interference power and transmit power, which is similar to the approach taken in [35]. The joint beamforming design problem is then formulated as

$$\max_{\mathbf{f}_{q,i}, \boldsymbol{\Theta}} \sum_{(q,i) \in \mathcal{P}} \|\mathbf{h}_{R,i} \boldsymbol{\Theta} \mathbf{H}_{q,R} \mathbf{f}_{q,i}\|_2^2 \quad (1a)$$

$$\text{s.t. } \|\mathbf{h}_{R,j} \boldsymbol{\Theta} \mathbf{H}_{q,R} \mathbf{f}_{q,i}\|_2^2 \leq I_{th}, \forall (q,i) \in \mathcal{P}, \forall j \neq i, \quad (1b)$$

$$|\boldsymbol{\Theta}_{mm}| \leq 1, m = 1, \dots, ML, \quad (1c)$$

$$\sum_i \|\mathbf{f}_{q,i}\|_2^2 \leq P^q, \forall q \in \mathcal{Q}, \quad (1d)$$

where the left hand side of (1b) corresponds to the pairwise instantaneous leakage interference power implied by the received signal model and I_{th} denotes the threshold for the pairwise instantaneous leakage interference power, which is usually obtained from long-term statistical measures or link budget to guarantee successful communication between the AP and UE [36]. P^q denotes the transmit power threshold and (1d) is the corresponding transmit power constraint for the q th AP. Notice that (1c) models the IRS as a passive device, i.e., there is no signal amplification and hence no additional power consumption.

III. GBD-BASED JOINT BEAMFORMING ALGORITHM

Notice from (1) that the composite channel between the q th AP and i th UE is written as $\mathbf{h}_{R,i} \boldsymbol{\Theta} \mathbf{H}_{q,R}$, where $\boldsymbol{\Theta}$ contains phase shift components. The problem is nonconvex but can be solved by applying the Generalized Benders decomposition [37], which is introduced in the following section. In short, (1) can be solved by splitting the problem into two subproblems, one solving for $\mathbf{f}_{q,i}$ and the other for $\boldsymbol{\Theta}$.

A. GBD REVISIT

GBD is a splitting algorithm devoted to solving problems of the form

$$\max_{\mathbf{x}, \mathbf{y}} f(\mathbf{x}, \mathbf{y}), \text{ s.t. } \mathbf{g}(\mathbf{x}, \mathbf{y}) \geq \mathbf{0}, \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}, \quad (2)$$

where $\mathbf{x} \in \mathbb{R}^{n_1}$ and $\mathbf{y} \in \mathbb{R}^{n_2}$ are variables with corresponding constraint sets \mathcal{X} and \mathcal{Y} , $f(\cdot, \cdot) : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ and $\mathbf{g}(\cdot, \cdot) : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightarrow \mathbb{R}^d$ denote the objective and constraint functions, respectively. \mathbf{y} is usually referred to as the complicating variable, i.e., if \mathbf{y} is fixed, then (2) is a convex problem and solution can be easily found. Obtaining solution for (2) is challenging as f and g are not necessarily convex with respect to \mathbf{x} and \mathbf{y} .

The GBD is an iterative algorithm, where a subproblem (SP) and relaxed master problem (RMP) is solved iteratively. An optimality cut is obtained by solving the SP and its corresponding optimal value serves as a lower bound for that of the original problem. The RMP is considered a relaxed version of the original problem, thus, its optimal value serves as an upper bound for that of the original problem. Convergence of the algorithm is achieved when the lower bound is equal to or greater than the upper bound. In practice, an ε -suboptimal solution is obtained when the gap between the lower and upper bound is within a predefined ε . Specifically, the SP is obtained by simply fixing $\mathbf{y} = \mathbf{y}^{(k)}$ in (2), resulting in the SP of the form $\max_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}^{(k)})$, s.t. $\mathbf{g}(\mathbf{x}, \mathbf{y}^{(k)}) \geq \mathbf{0} : \mathbf{u}$, $\mathbf{x} \in \mathcal{X}$, with \mathbf{u} being the Lagrange multiplier associated with $\mathbf{g}(\mathbf{x}, \mathbf{y}^{(k)}) \geq \mathbf{0}$. k denotes the GBD iteration. Optimal value of the SP is a lower bound for that of the original problem because \mathbf{x}^* has been obtained by fixing the other variable to a (nonoptimal) \mathbf{y}^k . $LB^{(k)} \triangleq \sup_{j \in \{1, \dots, k\}} f(\mathbf{x}^{(j)*}, \mathbf{y}^{(j)})$ denotes the greatest lower bound obtained among k number of GBD iterations.

The key idea behind the RMP is to approximate the projection of (2), which lies in the \mathbf{xy} -space, onto the \mathbf{y} -space with a series of approximations or cuts, that can be obtained after solving the SP. In other words, the goal is to create a problem in \mathbf{y} -space that yields the same optimal value and solution as the original (2) in the \mathbf{xy} -space. Assuming a feasible solution can be found for the SP, the RMP at the k th iteration is formulated as

$$\max_{\mathbf{y}, \gamma} \gamma \quad (3a)$$

$$\text{s.t. } \gamma \leq f(\mathbf{x}^{(\ell)}, \mathbf{y}) + \mathbf{u}^{(\ell)T} \mathbf{g}(\mathbf{x}^{(\ell)}, \mathbf{y}), \quad 1 \leq \ell \leq k, \quad (3b)$$

$$\mathbf{y} \in \mathcal{Y}, \quad (3c)$$

where constraints (3b) are optimality cuts. Since the RMP is a relaxed approximation of the original problem, γ^* is an upper bound of the optimal value of (2). Define $UB^{(k)} \triangleq \inf_{j \in \{1, \dots, k\}} \gamma^{(j)*}$ as the lowest upper bound obtained among k number of GBD iterations and note that $\gamma^{(j)*}$, for $j = 1$ to k is nonincreasing. The ε -suboptimal solution of (2) is obtained when $UB^{(k)} - LB^{(k)} \leq \varepsilon$. By sequentially applying cuts, the GBD is guaranteed to converge and its solution will converge to the solution of (2), which is optimal under certain conditions. In the case where the SP is not feasible, additional feasibility cuts need to be added to the RMP in the form of constraints. Further details about the feasibility cuts can be found in [37]. Optimality of the GBD scheme is described in the **Finite ε -convergence** theorem in [37] which states

Theorem 1 (Finite ε -Convergence [37]): *Assume that set of feasible \mathbf{x} and \mathbf{y} is a nonempty compact convex set and that objective and constraint functions are convex for each fixed \mathbf{x} . Further assume that the set of optimal Lagrangian dual multipliers \mathbf{u} is nonempty for any \mathbf{x} and is bounded. Then, for any given ε , the generalized Benders decomposition procedure terminates in a finite number of steps.*

The proof is given as a proof of [37, Th. 2.5].

B. SUBPROBLEM

Identifying Θ as the complicating variable, the SP is obtained by fixing Θ in (1) to be $\hat{\Theta}$, so that (2) transforms into

$$\begin{aligned} \max_{\mathbf{f}_{q,i}} \quad & \sum_{(q,i) \in \mathcal{P}} \|\mathbf{h}_{R,i} \hat{\Theta} \mathbf{H}_{q,R} \mathbf{f}_{q,i}\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{H}_{R,j} \hat{\Theta} \mathbf{H}_{q,R} \mathbf{f}_{q,i}\|_2^2 \leq I_{th}, \quad \forall (q,i) \in \mathcal{P}, \forall j \neq i, \\ & \sum_i \|\mathbf{f}_{q,i}\|_2^2 \leq P^q, \quad \forall q \in \mathcal{Q}. \end{aligned} \quad (4)$$

Even though (4) is nonconvex, this can be easily remedied by defining $\mathbf{W}_{q,i} \triangleq \mathbf{f}_{q,i} \mathbf{f}_{q,i}^H$ and applying semidefinite relaxation (SDR) to obtain

$$\max_{\mathbf{W}} \sum_{(q,i) \in \mathcal{P}} \text{tr}(\mathbf{h}_{R,i} \hat{\Theta} \mathbf{H}_{q,R} \mathbf{W}_{q,i} \mathbf{H}_{q,R}^H \hat{\Theta}^H \mathbf{h}_{R,i}^H) \quad (5a)$$

$$\begin{aligned} \text{s.t.} \quad & \text{tr}(\mathbf{h}_{R,j} \hat{\Theta} \mathbf{H}_{q,R} \mathbf{W}_{q,i} \mathbf{H}_{q,R}^H \hat{\Theta}^H \mathbf{h}_{R,j}^H) \leq I_{th}, \\ & \forall (q,i) \in \mathcal{P}, \forall j \neq i, \end{aligned} \quad (5b)$$

$$\mathbf{W}_{q,i} \geq 0, \quad \forall (q,i) \in \mathcal{P}, \quad (5c)$$

$$\sum_i \text{tr}(\mathbf{W}_{q,i}) \leq P^q, \quad \forall q \in \mathcal{Q}. \quad (5d)$$

$\mathbf{f}_{q,i}^*$ can be obtained from $\mathbf{W}_{q,i}^*$, the optimal solution of (5), by employing Gaussian randomization [7], [35], [40]. First, the eigenvalue decomposition of $\mathbf{W}_{q,i}^* = \mathbf{U}_{\mathbf{W}} \boldsymbol{\Sigma}_{\mathbf{W}} \mathbf{U}_{\mathbf{W}}^H$ is obtained, where $\mathbf{U}_{\mathbf{W}} \in \mathbb{C}^{N_T \times N_T}$ is a unitary matrix with all eigenvectors of $\mathbf{W}_{q,i}$, $\mathbf{v}_{\mathbf{W},1}, \dots, \mathbf{v}_{\mathbf{W},N_T}$, as column vectors, and $\boldsymbol{\Sigma}_{\mathbf{W}} = \text{Diag}(\lambda_{\mathbf{W},1}, \dots, \lambda_{\mathbf{W},N_T})$ contains all eigenvalues of $\mathbf{W}_{q,i}^*$ as diagonal elements arranged in descending order. C candidate rank-one solutions are generated next in the form of $\mathbf{f}_{q,i}^c = \mathbf{U}_{\mathbf{W}} \boldsymbol{\Sigma}_{\mathbf{W}}^{1/2} \mathbf{r}_c^{\mathbf{W}}$, where $\mathbf{r}_c^{\mathbf{W}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_T})$. Finally, the $\mathbf{f}_{q,i}^c$ that produces the largest objective value while satisfying all the constraints in (4) is chosen as the actual solution $\mathbf{f}_{q,i}^*$. It shall be noted that $\mathbf{W}_{q,i}^*$ is usually already rank-one, thus, $\mathbf{f}_{q,i}^* = \sqrt{\lambda_{\mathbf{W},1}} \mathbf{v}_{\mathbf{W},1}$. It is shown in the following section that $\mathbf{f}_{q,i}^*$ is not needed for the RMP and the rank-one approximation procedure is only needed after the convergence of the GBD algorithm instead of in each iteration.

Note that (5) is always feasible with the trivial solution $\mathbf{W} = \mathbf{0}$, thus no feasibility cuts in the RMP are needed. Moreover, (5) is separable in terms of $\mathbf{W}_{q,i}$ for different q , which allows (5) to be solved locally at each AP and in parallel across all APs.

C. RELAXED MASTER PROBLEM

To obtain the RMP, $\mathbf{f}_{q,i}$ is fixed to be $\mathbf{f}_{q,i} = \hat{\mathbf{f}}_{q,i}$, where $\hat{\mathbf{f}}_{q,i}$ is the solution of the SP in (4). Rewriting $\mathbf{h}_{R,i} \Theta$ as $\boldsymbol{\theta}^T \text{Diag}(\mathbf{h}_{R,i})$, where $\boldsymbol{\theta} \triangleq \text{diag}(\Theta)$, (5) becomes

$$\begin{aligned} \max_{\boldsymbol{\theta}} \quad & \sum_{(q,i) \in \mathcal{P}} \|\boldsymbol{\theta}^T \text{Diag}(\mathbf{h}_{R,i}) \mathbf{H}_{q,R} \hat{\mathbf{f}}_{q,i}\|_2^2 \\ \text{s.t.} \quad & \|\boldsymbol{\theta}^T \text{Diag}(\mathbf{h}_{R,j}) \mathbf{H}_{q,R} \hat{\mathbf{f}}_{q,i}\|_2^2 \leq I_{th}, \quad \forall (q,i) \in \mathcal{P}, \forall j \neq i, \\ & \|[\boldsymbol{\theta}]_m\| \leq 1, \quad m = 1, \dots, ML. \end{aligned} \quad (6)$$

Next, define $\tilde{\Theta} \triangleq \boldsymbol{\theta} \boldsymbol{\theta}^H \in \mathbb{C}^{ML \times ML}$ and apply SDR on $\tilde{\Theta}$ to obtain

$$\begin{aligned} \max_{\tilde{\Theta}} \quad & \sum_{(q,i) \in \mathcal{P}} \text{tr}(\hat{\mathbf{f}}_{q,i}^H \mathbf{H}_{q,R}^H \text{Diag}(\mathbf{h}_{R,i}^*) \tilde{\Theta}^* \text{Diag}(\mathbf{h}_{R,i}) \mathbf{H}_{q,R} \hat{\mathbf{f}}_{q,i}) \\ \text{s.t.} \quad & \text{tr}(\hat{\mathbf{f}}_{q,i}^H \mathbf{H}_{q,R}^H \text{Diag}(\mathbf{h}_{R,j}^*) \tilde{\Theta}^* \text{Diag}(\mathbf{h}_{R,j}) \mathbf{H}_{q,R} \hat{\mathbf{f}}_{q,i}) \leq I_{th}, \\ & \forall (q,i) \in \mathcal{P}, \forall j \neq i, \\ & \|[\tilde{\Theta}]_{m_1 m_2}\| \leq 1, \quad m_1, m_2 = 1, \dots, ML, \quad \tilde{\Theta} \succeq \mathbf{0}. \end{aligned} \quad (7)$$

Next step is to derive the Lagrangian function of (7). Define $\hat{\mathbf{W}}_{q,i} \triangleq \hat{\mathbf{f}}_{q,i} \hat{\mathbf{f}}_{q,i}^H$ as the primal solution and $\hat{\mathbf{u}} \in \mathbb{R}^{I(I-1)}$ as the dual solution of (5) with respect to only the pairwise instantaneous leakage power constraints and let $\hat{\mathbf{W}} = \{\hat{\mathbf{W}}_{q,i} | \forall (q,i) \in \mathcal{P}\}$. The Lagrangian function of (7) becomes $\mathcal{L}(\hat{\mathbf{W}}, \tilde{\Theta}, \hat{\mathbf{u}}) = \sum_{(q,i) \in \mathcal{P}} \text{tr}(\hat{\mathbf{W}}_{q,i}^H \mathbf{H}_{q,R}^H \text{Diag}(\mathbf{h}_{R,i}^*) \tilde{\Theta}^* \text{Diag}(\mathbf{h}_{R,i}) \mathbf{H}_{q,R}) + \hat{\mathbf{u}}^T (I_{th} \mathbf{1}_{I(I-1) \times 1} - \mathbf{g}(\hat{\mathbf{W}}, \tilde{\Theta}))$, where $\mathbf{g}(\hat{\mathbf{W}}, \tilde{\Theta}): (N_T \times N_T) \times (ML \times ML) \rightarrow I(I-1)$ contains all of the

pairwise instantaneous leakage interference power tuples. Specifically,

$$[\mathbf{g}(\widehat{\mathbf{W}}, \tilde{\Theta})]_{p(q,i,j)} = g(q, i, j) \triangleq \text{tr}\left(\widehat{\mathbf{W}}_{q,i} \mathbf{H}_{q,R}^H \text{Diag}(\mathbf{h}_{R,j}^*) \tilde{\Theta}^* \text{Diag}(\mathbf{h}_{R,j}) \mathbf{H}_{q,R}\right)$$

with $p(\cdot)$ being a one-to-one mapping from $\{(q, i, j) | (q, i) \in \mathcal{P}, (q, j) \in \mathcal{P}, i \neq j\}$ to $\{1, \dots, I(I-1)\}$. For example, if $Q = 2, I = 3$ and $\mathcal{P} = \{(1, 1), (1, 2), (2, 3)\}$. Then the possible (q, i, j) tuples are

$$(1, 1, 2) (1, 2, 1) (2, 3, 1) \\ (1, 1, 3) (1, 2, 3) (2, 3, 2)$$

where each column denotes the AP and user indices that generate and perceive the leakage interference. For instance, the first column indicates that AP 1 is transmitting to UE 1 and the leakage interference is perceived by UE 2 and 3. Similarly, the third column indicates that AP 2 is transmitting to UE 3, and the leakage interference is perceived by UE 1 and 2. In that case, $\mathbf{g}(\widehat{\mathbf{W}}, \tilde{\Theta}) \triangleq [g(1, 1, 2) \ g(1, 1, 3) \ \dots \ g(2, 4, 2) \ g(2, 4, 3)]^T \in \mathbb{R}^{I(I-1)}$.

Finally, the relaxed master problem at the k th iteration can be written as:

$$\max_{\tilde{\Theta}, z} z \quad (8a)$$

$$\text{s.t. } z \leq \mathcal{L}(\widehat{\mathbf{W}}^j, \tilde{\Theta}, \hat{\mathbf{u}}^j), j = 1, \dots, k \quad (8b)$$

$$|[\tilde{\Theta}]_{m_1 m_2}| \leq 1, \quad m_1, m_2 = 1, \dots, ML, \quad (8c)$$

$$\tilde{\Theta} \succeq \mathbf{0}, \quad (8d)$$

where $\widehat{\mathbf{W}}^j$ and $\hat{\mathbf{u}}^j$ are respectively the primal solution and the dual solution computed by solving the SP at the j th GBD iteration. Therefore, as indicated before, there is no need to invoke randomization to obtain $\hat{\mathbf{F}}^j$ from $\widehat{\mathbf{W}}^j$ until the last GBD iteration since only $\widehat{\mathbf{W}}^j$ is needed in the RMP. Similar to the randomization described above, θ^* can be obtained from $\tilde{\Theta}$ by employing Gaussian randomization. C candidates of θ can be generated with $\theta_c = \mathbf{U}_\theta \Sigma_\theta^{1/2} \mathbf{r}_c^\theta$, where $\mathbf{U}_\theta \in \mathbb{C}^{ML \times ML}$ and $\Sigma_\theta \in \mathbb{C}^{ML \times ML}$ are obtained from the eigendecomposition of $\tilde{\Theta} = \mathbf{U}_\theta \Sigma_\theta \mathbf{U}_\theta^H$, with $\mathbf{r}_c^\theta \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{ML})$. All θ_c 's are then normalized by the largest $|[\theta_c]_m|$ of its elements, so that all θ_c satisfy (8c). Finally, the θ_c that achieves the largest objective value while satisfying all the constraints in (8) is chosen as θ^* . Similar to $\mathbf{W}_{q,i}^*$, $\tilde{\Theta}^*$ is usually already of rank-one, thus, θ^* equals to the square root of the maximum eigenvalue multiplied by the maximum eigenvector of $\tilde{\Theta}^*$. In the rare occasion where (8c) is violated by such a θ^* , a feasible solution can be obtained by normalizing θ^* with its largest element. Different from $\mathbf{f}_{q,i}^*$, θ^* must be obtained at every iteration as the SP requires its value. Unlike $\mathbf{W}_{q,i}$ in the subproblem, (8) is not separable with respect to $\tilde{\Theta}$, so the phase shifters at different IRSs cannot be solved locally at each IRS, even if a hybrid IRS is used.

The proposed GBD based algorithm for joint beamformer design with given AP-UE pairing is summarized

Algorithm 1: GBD-Based Joint IRS-AP Design Algorithm With Given AP-UE Pairing (Single IRS)

- 1 Initialize ε , set $k = 1$, and randomly initialize θ ;
 - 2 One AP is assigned as the central access point (CAP);
 - 3 Channel information $\mathbf{H}_{q,R}, \mathbf{h}_{R,i}, q \in \mathcal{Q}, i \in \mathcal{I}$ is broadcast to each AP.
 - 4 **repeat**
 - 5 The CAP sends θ to all APs ;
 - 6 For each AP solve (5) and send $\mathbf{W}_{q,i}, \mathbf{u}^k$ and lower bound (LB) to the CAP ;
 - 7 (8) is solved at the CAP for $\tilde{\Theta}$ and upper bound (UB);
 - 8 Recover θ^* from $\tilde{\Theta}^*$ with randomization or selecting its maximum eigenvector ;
 - 9 Update $k = k + 1$;
 - 10 **until** $UB - LB < \varepsilon$;
 - 11 Each AP recovers $\mathbf{f}_{q,i}$ from $\mathbf{W}_{q,i}$ with randomization or selecting its maximum eigenvector;
-

in Algorithm 1. It is assumed there is a central unit that solves (8) for $\tilde{\Theta}$ and the upper bound (UB).

IV. GBD-BASED ALGORITHM WITH AP-UE PAIRING (GBD-MISDP)

Until this point, it was assumed that AP-UE pairing was performed *a priori* to the joint beamformer design, i.e., \mathcal{P} is known. In this section, the AP-UE pairing problem is incorporated into the formulation to assign UEs to APs in a sophisticated way.

A. PROBLEM FORMULATION

Let $v_{q,i} \in \{0, 1\}$ be a variable to represent pairing between the q th AP and i th UE. Specifically, $v_{q,i} = 1$ when the i th UE is served by the q th AP, and $v_{q,i} = 0$ otherwise. The new problem can then be formulated as

$$\max_{\mathbf{f}_{q,i}, \Theta, v_{q,i}} \sum_q \sum_i \|\mathbf{H}_{R,i} \Theta \mathbf{H}_{q,R} \mathbf{f}_{q,i}\|_2^2 \quad (9a)$$

$$\text{s.t. } \|\mathbf{H}_{R,j} \Theta \mathbf{H}_{q,R} \mathbf{f}_{q,i}\|_2^2 \leq I_{th}, \forall j \neq i, \forall q \quad (9b)$$

$$|[\Theta]_{mm}| \leq 1, m = 1, \dots, ML, \quad (9c)$$

$$v_{q,i} \in \{0, 1\}, \forall q, \forall i \quad (9d)$$

$$\sum_i \|\mathbf{f}_{q,i}\|_2^2 \leq P^q, \forall q, \quad (9e)$$

$$\sum_q v_{q,i} = 1, \forall i \quad (9f)$$

$$\|\mathbf{f}_{q,i}\|_2^2 \leq v_{q,i} P^q, \forall q, \forall i. \quad (9g)$$

Note that \mathcal{P} now has been implicitly enumerated to include all possible AP-UE pairs in (9a) and (9b) by summing every AP and UE. Eq. (9f) ensures each UE is served by a single AP. Eq. (9g) is the big- M reformulation of the logic constraint “if $v_{q,i}$ equals to 0, then $\|\mathbf{f}_{q,i}\|_2^2$ equals 0”, which links $v_{q,i}$ and $\mathbf{f}_{q,i}$. P^q is chosen as the big- M such that when

$v_{q,i} = 1$, (9g) becomes $\|\mathbf{f}_{q,i}\|_2^2 \leq P^q$ and always holds due to (9e). Similar to the original formulation in (2), (9) is nonconvex, therefore, a similar strategy as the one described in Section III can be used to solve (9). The subproblem and associated relaxed master problem are derived in the following sections.

B. SUBPROBLEM

Identifying Θ as the complicating variable, the subproblem is obtained by fixing Θ in (9). For fixed $\Theta = \hat{\Theta}$, the subproblem of (9) is

$$\begin{aligned} \max_{\mathbf{f}_{q,i}, v_{q,i}} \quad & \sum_q \sum_i \|\mathbf{h}_{R,i} \hat{\Theta} \mathbf{H}_{q,R} \mathbf{f}_{q,i}\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{h}_{R,j} \hat{\Theta} \mathbf{H}_{q,R} \mathbf{f}_{q,i}\|_2^2 \leq I_{th}, \forall j \neq i, \forall q \in \mathcal{Q}, \\ & v_{q,i} \in \{0, 1\}, \forall q \in \mathcal{Q}, \forall i \in \mathcal{I}, \\ & \sum_i \|\mathbf{f}_{q,i}\|_2^2 \leq P_{th}, \forall q \in \mathcal{Q}, \\ & \sum_q v_{q,i} = 1, \forall i \in \mathcal{I}, \\ & \|\mathbf{f}_{q,i}\|_2^2 \leq v_{q,i} P_{th}, \forall q \in \mathcal{Q}, \forall i \in \mathcal{I}. \end{aligned} \quad (10)$$

Notice again that (10) is nonconvex. Similar to Section III-B, define $\mathbf{W}_{q,i} \triangleq \mathbf{f}_{q,i} \mathbf{f}_{q,i}^H$ and applying SDR, then (10) can be transformed to the MISDP problem

$$\begin{aligned} \max_{\mathbf{W}, v_{q,i}} \quad & \sum_{(q,i) \in \mathcal{P}} \text{tr}(\mathbf{h}_{R,i} \hat{\Theta} \mathbf{H}_{q,R} \mathbf{W}_{q,i} \mathbf{H}_{q,R}^H \hat{\Theta}^H \mathbf{h}_{R,i}^H) \quad (11a) \\ \text{s.t.} \quad & \text{tr}(\mathbf{h}_{R,j} \hat{\Theta} \mathbf{H}_{q,R} \mathbf{W}_{q,i} \mathbf{H}_{q,R}^H \hat{\Theta}^H \mathbf{h}_{R,j}^H) \leq I_{th}, \\ & \quad \quad \quad \forall q \in \mathcal{Q}, \forall j \neq i, \quad (11b) \\ & v_{q,i} \in \{0, 1\}, \forall q \in \mathcal{Q}, \forall i \in \mathcal{I}, \quad (11c) \\ & \mathbf{W}_{q,i} \succeq 0, \forall q \in \mathcal{Q}, \forall i \in \mathcal{I}, \quad (11d) \\ & \sum_i \text{tr}(\mathbf{W}_{q,i}) \leq P^q, \forall q \in \mathcal{Q}, \quad (11e) \\ & \sum_q v_{q,i} = 1, \forall i \in \mathcal{I}, \quad (11f) \\ & \text{tr}(\mathbf{W}_{q,i}) \leq v_{q,i} P^q, \forall q \in \mathcal{Q}, \forall i \in \mathcal{I}, \quad (11g) \end{aligned}$$

which requires a MISDP solver to find the global optimal solution, that can incur high computational and time complexity. Note that (11) is always feasible with the trivial solution $\mathbf{W} = \mathbf{0}$, thus no feasibility cuts in the RMP are needed. Unlike (5), solution for (11) cannot be attained at each AP in parallel due to the coupling in (11f). Similar to (5) in Section III-B, it is not necessary to recover the rank-one solutions $\mathbf{f}_{q,i}^*$ from $\mathbf{W}_{q,i}^*$ until after the convergence of the GBD algorithm. When $v_{q,i}^* = 0$ and $\text{tr}(\mathbf{W}_{q,i}^*) = 0$, it is easy to see that $\mathbf{f}_{q,i}^* = \mathbf{0}$. In the other case when $v_{q,i} = 1$, note that (11) and (5) have the same $\mathbf{W}_{q,i}$ related constraints except for (11g), which is guaranteed to hold if (11e) is satisfied. Thus, the solutions obtained from the same Gaussian randomization or the maximum eigenvector method are feasible solutions to (11) and can be used to recover $\mathbf{f}_{q,i}^*$ from $\mathbf{W}_{q,i}^*$

Algorithm 2: GBD-MISDP Based Algorithm or GBD-Based Algorithm With Heuristic AP-UE Pairing (Single IRS)

```

1 Initialize  $\varepsilon$ , set  $k = 1$ , and randomly initialize  $\theta$ ;
2 One AP is assigned as the central access point (CAP);
3  $\mathbf{H}_{q,R}, \mathbf{h}_{R,i}, q \in \mathcal{Q}, i \in \mathcal{I}$  are estimated and sent to CAP if MISDP or to all APs if heuristic pairing.
4 repeat
5   if GBD-MISDP then
6     At the CAP, solve (11) using the MISDP algorithm and obtain the LP ;
7   else
8     Run the ILR heuristic AP-UE pairing subroutine;
9     Each AP solve (5) and send  $\mathbf{W}_{q,i}, \mathbf{u}^k$  and the associated lower bound (LB) to the CAP ;
10  end
11  The CAP solves (8) for  $\tilde{\Theta}$  and calculates UB ;
12  Recover  $\theta$  from  $\tilde{\Theta}$  with randomization or selecting its maximum eigenvector;
13  Update  $k = k + 1$ 
14 until  $UB - LB < \varepsilon$ ;
15 if GBD-MISDP then
16   At the CAP, recover  $\mathbf{f}_{q,i}$  from  $\mathbf{W}_{q,i}$  with randomization or selecting its maximum eigenvector and send the designed beamformer to all APs.
17 else
18   Each AP recovers  $\mathbf{f}_{q,i}$  from  $\mathbf{W}_{q,i}$  with randomization or selecting its maximum eigenvector ;
19 end

```

C. RELAXED MASTER PROBLEM

The RMP is obtained by following the same procedure as in Section III-C, which results in exactly the same problem as (6), with $\mathcal{P} \triangleq \{(q, i) | v_{q,i} = 1\}$ and $\hat{v}_{q,i}$ is the solution of (11). The same procedure for recovering rank-one solutions θ^* from $\tilde{\Theta}^*$ in Section III-C is also used. The proposed GBD-MISDP algorithm for single IRS is summarized in Algorithm 2. In a multi-IRS scenario, the CAP can also be used to find the phase shifters for different IRSs.

V. HEURISTIC AP-UE SUBROUTINES

Although the GBD-MISDP algorithm can solve (9), the use of MISDP solver in solving the subproblem in (11) incur large computational and time complexity as will be shown in Section VII. A heuristic algorithms based on transmit power called iterative link removal (ILR), is proposed in the following to find a suboptimal AP-UE pairing.

The ILR is called by the GBD-ILR algorithm in Algorithm 2, which first assumes all channel state information and θ are sent to the APs. The ILR then iteratively designs $\mathbf{W}_{q,i}$, and remove the AP-UE link with the

Algorithm 3: AP-UE Pairing Subroutine: Iterative Link Removal

```

1 Initialize:  $v_{q,i} = 1, \forall (q, i), B = I$ 
2 repeat
3   Each AP solves (5) with  $\mathcal{P} \triangleq \{(q, i) | v_{q,i} = 1\}$  to
   get  $\mathbf{W}_{q,i}$ 
4   Reduce the number of  $B$ 
5   if  $B \geq 4\lceil I/Q \rceil$  then
6      $B = B/2$ ;
7   else if  $B \leq 2\lceil I/Q \rceil$  then
8      $B = \lceil \gamma I/Q \rceil$ ;
9   else
10     $B = B - \lceil 0.5I/Q \rceil$ ;
11  Each AP computes  $tr(\mathbf{W}_{q,i})$  and broadcast the
  information. Each AP chooses to serve the top  $B$ 
  UEs with the largest resulting  $tr(\mathbf{W}_{q,i})$  and set all
  other  $v_{q,i} = 0$ .
12  for Each UE not served do
13    Each AP broadcast its ranking of that UE
14    The AP with the highest ranking serves the UE
    and drops an user, under the condition of not
    leaving the said user not served.
15 until  $B = \lceil \gamma I/Q \rceil$ ;
16 for Each UE served by multiple APs do
17  All APs except for the one that UE has the largest
   $tr(\mathbf{W}_{q,i})$  with drop the said UE, such that each UE
  is served by one AP.

```

smallest transmit power $tr(\mathbf{W}_{q,i})$, similar to [38]. With the same consideration to prevent the number of UE served at each AP being uneven, the iterative link removal algorithm can be described as follows. At each iteration, the SP in (5) is solved to obtain $\mathbf{W}_{q,i}$. A ranking of UEs based on $tr(\mathbf{W}_{q,i})$ is computed and each AP picks the top B UEs with the largest $tr(\mathbf{W}_{q,i})$ and turns off the other beamformers, where B again serves the purpose of preventing overloading any one AP. If any UE is not served, the AP with the highest transmit power based ranking with that UE would be chosen to serve it and drops another UE. B gradually decreases over iterations until the termination condition is met. Finally, if any UE is served by multiple APs, all APs except for the one that UE has the largest $tr(\mathbf{W}_{q,i})$ will drop the said UE. The overall procedure is described in Algorithm 3. Notice that γ is a hyperparameter to control the fairness among APs and has to be selected beforehand.

The GBD-ILR algorithm is summarized in Algorithm 2, with Algorithm 3 used in line 8. Since the SP solutions $\mathbf{u}, \mathbf{W}_{q,i}$, and $v_{q,i}$ are not guaranteed to be the optimal solutions of (11), the finite ε -convergence in [37] is replaced with a weaker statement that the optimal objective value produced by the RMP in (8) and the gap $UB - LB$ are nonincreasing.

In summary, it is important to point out that the GBD algorithm can only guarantee that local optimal solution can be found, which implies at least a local optimal solution for (1) and (9) can be attained using Algorithms 1 and 2, respectively. However, as stated in Theorem 1, both algorithms will converge in finite number of steps as both problems satisfy the premise stated in the theorem.

VI. SAG ALGORITHM FOR JOINT AIRS POSITIONING, PASSIVE AND ACTIVE BEAMFORMER DESIGN

In previous discussions, the locations of IRS panels were fixed and determined arbitrarily. To further enhance the overall performance of IRS-assisted system, the AIRS-assisted system, with a single AIRS (i.e., $L = 1$), and the associated positioning problem shall be considered hereafter. The proposed SAG algorithm uses the SCA, alternating direction method of multipliers (ADMM), and the proposed GBD based method in Section III to jointly solve the AIRS positioning, passive and active beamformer design problem which is described in the sequel. As before, Q denotes the total number of APs, each with N_T transmit antennas, I denotes the total number of single-antenna UEs, and M denotes the number of reflecting elements on the AIRS. It is assumed that the UEs have already been paired with the APs using the random pairing method assumed in Section III.

A. SPATIAL CORRELATION MODEL

Since the development of the AIRS depends on an analytical channel model, a complete description of the model shall be given first, followed by the derivation of the SAG. The channel between the APs and the AIRS, and between the AIRS and UEs are modeled as

$$\begin{aligned} \mathbf{H}_{q,R} &= \mathbf{H}_{q,R,w} \mathbf{P}_{q,R} \in \mathbb{C}^{M \times N_T} \quad \text{and} \\ \mathbf{h}_{R,i} &= \mathbf{h}_{R,i,w} \mathbf{P}_{R,i} \in \mathbb{C}^{1 \times M}, \end{aligned} \quad (12)$$

respectively. It is assumed that both channels contain LOS component, hence, $\mathbf{H}_{q,R,w}$ and $\mathbf{h}_{R,i,w}$ are complex Gaussian random matrix and vector with i.i.d. nonzero-mean μ and σ variance complex Gaussian entries, which model the micro-fading part of $\mathbf{H}_{q,R}$ and $\mathbf{H}_{R,i}$, respectively.

$$\mathbf{P}_{q,R} = \text{Diag} \left(\sqrt{d_{q,R,1}^{-\alpha} 10^{\xi_{q,1}/10}}, \dots, \sqrt{d_{q,R,N_T}^{-\alpha} 10^{\xi_{q,N_T}/10}} \right) \quad (13a)$$

$$\mathbf{P}_{R,i} = \text{Diag} \left(\sqrt{d_{R,i,1}^{-\alpha} 10^{\xi_{i,1}/10}}, \dots, \sqrt{d_{R,i,M}^{-\alpha} 10^{\xi_{i,M}/10}} \right) \quad (13b)$$

are $N_T \times N_T$ and $M \times M$ diagonal matrices which model the macro-fading part of their respective channels. α denotes the path loss exponent. $10^{\xi_{q,n}/10}$ and $10^{\xi_{i,m}/10}$, for $n = 1, \dots, N_T$ and $m = 1, \dots, M$, are independent of the path loss $d_{q,R,n}^{-\alpha}$ and $d_{R,i,m}^{-\alpha}$, respectively, and are log-normal distributed random variable, with $\xi_{q,n}$ and $\xi_{i,m}$ being normal distributed random variables with zero mean and variance equals to 4 dB. $d_{q,R,n}$ and $d_{R,i,m}$ are the distance from the n th antenna

of the q th AP to the AIRS and from the m th element of the AIRS to the i th UE, respectively.

Let ℓ , \mathbf{b}_q and $\mathbf{u}_i \in \mathbb{R}^3$ represent the coordinates of AIRS, APs and UEs. Assuming hereafter that the distances from AP to AIRS and from AIRS to UE are far enough compared to the inter-distance of antennas and passive elements on the AIRS, so that $d_{q,R,n} = d_{q,R} = \|\mathbf{b}_q - \ell\|_2$, for $n = 1, \dots, N_T$ and $d_{R,i,m} = d_{R,i} = \|\ell - \mathbf{u}_i\|_2$, for $m = 1, \dots, M$, $\xi_{q,n} = \xi_q$, for $n = 1, \dots, N_T$ and $\xi_{i,m} = \xi_i$, for $m = 1, \dots, M$. Then, (13) can be written as

$$\begin{aligned} \mathbf{P}_{q,R} &= \mathbf{I}_{N_T} \sqrt{10^{\xi_{q,R}/10} / \|\mathbf{b}_q - \ell\|_2^\alpha} \\ \mathbf{P}_{R,i} &= \mathbf{I}_M \sqrt{10^{\xi_{R,i}/10} / \|\ell - \mathbf{u}_i\|_2^\alpha} \end{aligned} \quad (14)$$

Assuming $f_{q,i}$, Θ and the AP-user pairings are known. Using the channel model in (12) with (14), ℓ can then be solved by

$$\begin{aligned} \max_{\ell} \quad & \sum_{(q,i) \in \mathcal{P}} \frac{10^{(\xi_{q,R} + \xi_{R,i})/10}}{(\|\ell - \mathbf{b}_q\| \|\ell - \mathbf{u}_i\|)^\alpha} \|\mathbf{h}_{R,i,w} \Theta \mathbf{H}_{q,R,w} \mathbf{f}_{q,i}\|_2^2 \\ \text{s.t.} \quad & \frac{10^{(\xi_{q,R} + \xi_{R,j})/10}}{(\|\ell - \mathbf{b}_q\| \|\ell - \mathbf{u}_j\|)^\alpha} \|\mathbf{h}_{R,j,w} \Theta \mathbf{H}_{q,R,w} \mathbf{f}_{q,i}\|_2^2 \leq I_{th}, \\ & \forall (q,i) \in \mathcal{P}, j \neq i. \end{aligned} \quad (15)$$

B. CONSTRAINT REFORMULATION VIA SCA

It is obvious that (15) is nonconvex in both the objective and constraints. Focusing on a particular (q,i) th AP-user pair and $j \neq i$, the constraint can be rewritten as

$$\begin{aligned} & 10^{\frac{\xi_{q,R} + \xi_{R,j}}{10}} \|\mathbf{h}_{R,j,w} \Theta \mathbf{H}_{q,R,w} \mathbf{f}_{q,i}\|_2^2 \leq I_{th} \\ & \quad (\|\ell - \mathbf{b}_q\| \|\ell - \mathbf{u}_j\|)^\alpha \\ \Leftrightarrow & 10^{\frac{\xi_{q,R} + \xi_{R,j}}{10\alpha}} |\mathbf{h}_{R,j,w} \Theta \mathbf{H}_{q,R,w} \mathbf{f}_{q,i}|^{2/\alpha} \leq I_{th}^{1/\alpha} \\ & \quad (\|\ell - \mathbf{b}_q\| \|\ell - \mathbf{u}_j\|). \end{aligned}$$

The norm product on the right hand side can be replaced by its lower bound $(\ell - \mathbf{b}_q)^T (\ell - \mathbf{u}_j)$ resulting in

$$10^{\frac{\xi_{q,R} + \xi_{R,j}}{10\alpha}} |\mathbf{h}_{R,j,w} \Theta \mathbf{H}_{q,R,w} \mathbf{f}_{q,i}|^{2/\alpha} \leq I_{th}^{1/\alpha} \left| (\ell - \mathbf{b}_q)^T (\ell - \mathbf{u}_j) \right|.$$

Define $f_{c,q,i,j}(\ell) \triangleq (\ell - \mathbf{b}_q)^T (\ell - \mathbf{u}_j)$. Since $f_{c,q,i,j}(\ell)$ is convex, to make the constraint convex, it shall be linearized as $f_{c,q,i,j}(\ell_0) + \nabla f_{c,q,i,j}(\ell_0)^T (\ell - \ell_0)$, where ℓ_0 is a fixed point, so that right hand side of the constraint can be written as

$$I_{th}^{1/\alpha} |f_{c,q,i,j}(\ell)| \approx I_{th}^{1/\alpha} \left| f_{c,q,i,j}(\ell_0) + \nabla f_{c,q,i,j}(\ell_0)^T (\ell - \ell_0) \right|,$$

where $\nabla f_{c,q,i,j}(\ell_0) = 2\ell_0 - \mathbf{b}_q - \mathbf{u}_j \in \mathbb{R}^3$. Thus the constraints can be written as

$$10^{\frac{\xi_{q,R} + \xi_{R,j}}{10\alpha}} |\mathbf{h}_{R,j,w} \Theta \mathbf{H}_{q,R,w} \mathbf{f}_{q,i}|^{2/\alpha} \leq I_{th}^{1/\alpha} \left| f_{c,q,i,j}(\ell_0) + \nabla f_{c,q,i,j}(\ell_0)^T (\ell - \ell_0) \right|$$

$$\begin{aligned} \Leftrightarrow c_{q,i,j} &\leq I_{th}^{1/\alpha} \left| \nabla f_{c,q,i,j}(\ell_0)^T \ell - (\|\ell_0\|_2^2 - \mathbf{b}_q^T \mathbf{u}_j) \right| \\ \Leftrightarrow c_{q,i,j} &\leq \left| I_{th}^{1/\alpha} \nabla f_{c,q,i,j}(\ell_0)^T \ell - b_{q,i,j} \right|, \end{aligned}$$

$\forall (q,i) \in \mathcal{P}, j \neq i$, where $c_{q,i,j} \triangleq 10^{\frac{\xi_{q,R} + \xi_{R,j}}{10\alpha}} |\mathbf{h}_{R,j,w} \Theta \mathbf{H}_{q,R,w} \mathbf{f}_{q,i}|^{2/\alpha}$, and $b_{q,i,j} \triangleq I_{th}^{1/\alpha} (\|\ell_0\|_2^2 - \mathbf{b}_q^T \mathbf{u}_j)$. To account for all $(i,j) \in \mathcal{P}, \forall j \neq i$, and $\forall q \in \mathcal{Q}$, define $\mathbf{c} \triangleq [c_{1,1,2} \ c_{1,1,3} \ \dots \ c_{Q,I,1} \ \dots \ c_{Q,I,I-1}]^T \in \mathbb{R}^{I(I-1)}$,

$$\mathbf{F} \triangleq I_{th}^{1/\alpha} \begin{bmatrix} \nabla f_{c,1,1,2}(\ell_0)^T \\ \nabla f_{c,1,1,3}(\ell_0)^T \\ \vdots \\ \nabla f_{c,Q,I,1}(\ell_0)^T \\ \vdots \\ \nabla f_{c,Q,I,I-1}(\ell_0)^T \end{bmatrix} \in \mathbb{R}^{I(I-1) \times 3},$$

and $\mathbf{b} \triangleq [b_{1,1,2} \ b_{1,1,3} \ \dots \ b_{Q,I,1} \ \dots \ b_{Q,I,I-1}]^T \in \mathbb{R}^{I(I-1)}$. Then the constraints can be written as

$$\mathbf{c} \leq |\mathbf{F}\ell - \mathbf{b}| \Leftrightarrow \mathbf{c} \leq \mathbf{F}\ell - \mathbf{b} \text{ and } \mathbf{c} \leq -\mathbf{F}\ell + \mathbf{b}.$$

Since only one of the two constraints above needs to be included in the problem, to guarantee this, a ‘‘big N ’’ term is added to the constraints. Define a binary variable $a_{q,i,j} \in \{0, 1\}$, and the corresponding vector $\mathbf{a} \triangleq [a_{1,1,2} \ a_{1,1,3} \ \dots \ a_{Q,I,1} \ \dots \ a_{Q,I,I-1}]^T \in \{0, 1\}^{I(I-1)}$. Then

$$\mathbf{c} < eq\mathbf{F}\ell - \mathbf{b} + N\mathbf{a} \quad (16)$$

$$\mathbf{c} < eq - \mathbf{F}\ell + \mathbf{b} + N(\mathbf{1}_{I(I-1) \times 1} - \mathbf{a}), \quad (17)$$

where N is a ‘‘big’’ value. To clarify, suppose $[\mathbf{a}]_i = 0$, then $[\mathbf{c}]_i$ in (17) is guaranteed to be less than the right hand side of (17), so that the only constraint that needs to be included in the problem is $[\mathbf{c}]_i \leq [\mathbf{F}\ell - \mathbf{b}]_i$, which corresponds to the i th element of (16). Conversely, if $[\mathbf{a}]_i = 1$, then (16) is guaranteed to be satisfied so there is no need to include it in the problem; only (17) needs to be explicitly included.

C. ADMM-BASED ALGORITHM

To solve the positioning problem, the constraints are further modified by adding a nonnegative slack variable $\mathbf{s}_0, \mathbf{s}_1 \geq \mathbf{0}_{I(I-1) \times 1}$ so they become

$$\mathbf{c} + \mathbf{s}_0 = \mathbf{F}\ell - \mathbf{b} + N\mathbf{a}$$

$$\mathbf{c} + \mathbf{s}_1 = -\mathbf{F}\ell + \mathbf{b} + N(\mathbf{1}_{I(I-1) \times 1} - \mathbf{a})$$

$$\mathbf{s}_0, \mathbf{s}_1 \succ eq\mathbf{0}_{I(I-1) \times 1}.$$

Define $f_0(\ell)$ as the objective in (15), then the design problem for the AIRS position can be written as

$$\max_{\ell, \mathbf{a}, \mathbf{s}} f_0(\ell)$$

$$\text{s.t. } \mathbf{c} + \mathbf{s}_0 = \mathbf{F}\ell - \mathbf{b} + N\mathbf{a}$$

$$\mathbf{c} + \mathbf{s}_1 = -\mathbf{F}\ell + \mathbf{b} + N(\mathbf{1}_{I(I-1) \times 1} - \mathbf{a})$$

$$\mathbf{s}_0, \mathbf{s}_1 \geq \mathbf{0}_{I(I-1) \times 1}. \quad (18)$$

This can now be solved using the ADMM, where (18) will be solved in 4 steps, in which each of the variables in (18)

is updated individually, in addition to a dual variable. The augmented Lagrangian function is written as

$$\begin{aligned} \mathcal{L}_\rho(\ell, \mathbf{s}_0, \mathbf{s}_1, \lambda_0, \lambda_1) &= f_0(\ell) - \frac{\rho_0}{2} \|\mathbf{F}\ell - \mathbf{b} + N\mathbf{a} - \mathbf{c} - \mathbf{s}_0\|_2^2 \\ &\quad - \frac{\rho_1}{2} \|\mathbf{F}\ell + \mathbf{b} + N(\mathbf{1}_{I(I-1) \times 1} - \mathbf{a}) - \mathbf{c} - \mathbf{s}_1\|_2^2 \\ &\quad + \lambda_0^T (\mathbf{F}\ell - \mathbf{b} + N\mathbf{a} - \mathbf{c} - \mathbf{s}_0) \\ &\quad + \lambda_1^T (-\mathbf{F}\ell + \mathbf{b} + N(\mathbf{1}_{I(I-1) \times 1} - \mathbf{a}) - \mathbf{c} - \mathbf{s}_1) \\ &= f_0(\ell) - \frac{\rho_0}{2} \|\mathbf{F}\ell - \mathbf{b} + N\mathbf{a} - \mathbf{c} - \mathbf{s}_0 - \lambda_{s,0}\|_2^2 \\ &\quad - \frac{\rho_1}{2} \|\mathbf{F}\ell + \mathbf{b} + N(\mathbf{1}_{I(I-1) \times 1} - \mathbf{a}) - \mathbf{c} - \mathbf{s}_1 - \lambda_{s,1}\|_2^2, \end{aligned}$$

where $\lambda_{s,i} \triangleq \frac{\lambda_i}{\rho_i}$, for $i = 1, 2$, denotes the scaled dual variable. The second equality is the scaled form of the augmented Lagrangian function. Since either (16) or (17) needs to be explicitly included in the problem formulation, that implies

$$[\mathbf{a}]_i = \begin{cases} 0, & [\mathbf{F}\ell - \mathbf{b}]_i \geq [\mathbf{c}]_i \\ 1, & [\mathbf{F}\ell - \mathbf{b}]_i \leq [\mathbf{c}]_i. \end{cases}$$

Since the elements of \mathbf{c} is nonnegative, then the above can be rewritten as

$$[\mathbf{a}]_i = \begin{cases} 0, & [\mathbf{F}\ell - \mathbf{b}]_i \geq 0 \\ 1, & [\mathbf{F}\ell - \mathbf{b}]_i \leq 0, \end{cases} \quad (19)$$

for $i = 1, \dots, I(I-1)$. Hence, the augmented Lagrangian can be modified as

$$\mathcal{L}_\rho(\ell, \mathbf{s}, \lambda_s) = f_0(\ell) - \frac{\rho}{2} \|\mathbf{p}(\ell, \mathbf{s}, \lambda_s)\|_2^2, \quad (20)$$

where $[\mathbf{p}(\ell, \mathbf{s}, \lambda_s)]_i = \begin{cases} [\mathbf{F}\ell - \mathbf{b} - \mathbf{c} - \mathbf{s} - \lambda_s]_i, & [\mathbf{a}]_i = 0 \\ [-\mathbf{F}\ell + \mathbf{b} - \mathbf{c} - \mathbf{s} - \lambda_s]_i, & [\mathbf{a}]_i = 1 \end{cases}$. In other words, only one ρ , slack variable, $\mathbf{s} \in \mathbb{C}^{(I-1)I \times 1}$ define as $[\mathbf{s}]_i = \begin{cases} [\mathbf{s}_0]_i, & [\mathbf{a}]_i = 0 \\ [\mathbf{s}_1]_i, & [\mathbf{a}]_i = 1 \end{cases}$, and scaled dual variable are needed. Therefore, the location of the AIRS, ℓ , can be obtained by solving

$$\ell^{k+1} = \arg \max_{\ell} f_0(\ell) - \frac{\rho^k}{2} \|\mathbf{p}(\ell, \mathbf{s}^k, \lambda_s^k)\|_2^2, \quad (21)$$

where k is the iteration index of the ADMM and the variables \mathbf{s} and λ_s in $\mathbf{p}(\ell, \mathbf{s}, \lambda_s)$ has now been replaced by their k th ADMM iterate. Notice that (21) is a nonconvex problem and it is solved using the gradient ascent with momentum (GAM) method and its motivation will be shown in Section VII. Let $\mathbf{g}(i) \triangleq \nabla_{\ell} (f_0(\ell) - \frac{\rho^k}{2} \|\mathbf{p}(\ell, \mathbf{s}^k, \lambda_s^k)\|_2^2)$ be the gradient of the objective in (21) at the i th iteration of the GAM. Then ℓ shall be updated as

$$\begin{aligned} \mathbf{v}(i) &= \begin{cases} \mathbf{g}(i), & i = 0 \\ \beta \mathbf{v}(i-1) + \mathbf{g}(i), & i > 0 \end{cases} \\ \ell(i+1) &= \ell(i) + \gamma \mathbf{v}(i), \end{aligned}$$

where β and γ are the momentum factor and step size of the GAM, respectively. When the GAM converges, then $\ell^{k+1} = \ell(i+1)$. After ℓ^{k+1} is obtained, it is then substituted

into (19) for ℓ to find the i th element of \mathbf{a} , $\forall i$, resulting in \mathbf{a}^{k+1} .

Since $f_0(\ell)$ is not a function of the slack variable \mathbf{s} , solving for \mathbf{s} only involves the quadratic penalty term in (20), which can be obtained by

$$\mathbf{s}^{k+1} = \arg \max_{\mathbf{s} \geq \mathbf{0}} - \|\mathbf{p}_s(\ell^{k+1}, \mathbf{s}, \lambda_s^k)\|_2^2, \quad (22)$$

where

$$[\mathbf{p}_s \ell, \mathbf{s}, \lambda_s]_i \triangleq \begin{cases} [\mathbf{F}\ell - \mathbf{b} - \mathbf{c} - \mathbf{s} - \lambda_s]_i, & [\mathbf{a}^{k+1}]_i = 0, \\ [-\mathbf{F}\ell + \mathbf{b} - \mathbf{c} - \mathbf{s} - \lambda_s]_i, & [\mathbf{a}^{k+1}]_i = 1 \end{cases}$$

Finally, λ_s by first defining

$$\mathbf{p}_\lambda(\ell^{k+1}, \mathbf{s}^{k+1}) \triangleq \begin{cases} [\mathbf{F}\ell^{k+1} - \mathbf{b} - \mathbf{c} - \mathbf{s}^{k+1}]_i, & [\mathbf{a}^{k+1}]_i = 0 \\ [-\mathbf{F}\ell^{k+1} + \mathbf{b} - \mathbf{c} - \mathbf{s}^{k+1}]_i, & [\mathbf{a}^{k+1}]_i = 1. \end{cases}$$

Then λ_s can then be found by

$$\lambda_s^{k+1} = \lambda_s^k - \mathbf{p}_\lambda(\ell^{k+1}, \mathbf{s}^{k+1}). \quad (23)$$

Using, (21), (19), (22), and (23), then the ADMM algorithm for finding ℓ becomes

$$\ell - \text{step: } \ell^{k+1} = \arg \max_{\ell} f_0(\ell) - \frac{\rho^k}{2} \|\mathbf{p}(\ell, \mathbf{s}^k, \lambda_s^k)\|_2^2 \quad (24a)$$

$$\mathbf{a} - \text{step: } [\mathbf{a}^{k+1}]_i = \begin{cases} 0, & [\mathbf{F}\ell^{k+1} - \mathbf{b}]_i \geq 0 \\ 1, & [\mathbf{F}\ell^{k+1} - \mathbf{b}]_i \leq 0 \end{cases} \quad (24b)$$

$$\mathbf{s} - \text{step: } \mathbf{s}^{k+1} = \arg \max_{\mathbf{s} \geq \mathbf{0}} - \|\mathbf{p}_s(\ell^{k+1}, \mathbf{s}, \lambda_s^k)\|_2^2 \quad (24c)$$

$$\lambda_s - \text{step: } \lambda_s^{k+1} = \lambda_s^k - \mathbf{p}_\lambda(\ell^{k+1}, \mathbf{s}^{k+1}). \quad (24d)$$

ρ^k is updated as

$$\rho^{k+1} = \begin{cases} \tau \rho^k, & \|\mathbf{r}^k\|_2 > \mu \|\mathbf{d}^k\|_2 \\ \rho^k / \tau, & \|\mathbf{r}^k\|_2 < \mu \|\mathbf{d}^k\|_2 \\ \rho^k, & \text{otherwise,} \end{cases} \quad (25)$$

where $\tau = 1.05$ and $\mu = 10$, with $\rho^0 = 1$. $\mathbf{r}^k \triangleq \mathbf{F}\ell^{k+1} - \mathbf{b} - \mathbf{s}^{k+1} - \mathbf{c}$ denotes the primal residual and $\mathbf{d}^k \triangleq \rho^k (\mathbf{F}(\mathbf{s}^k - \mathbf{s}^{k+1}))$ denotes the dual residual at the k th iteration. The ADMM portion of the SAG will terminate when $\|\mathbf{r}^k\|_2 \leq \varepsilon_{pri}^k$ and $\|\mathbf{d}^k\|_2 \leq \varepsilon_{dual}^k$, which are defined as

$$\begin{aligned} \varepsilon_{pri}^k &= \sqrt{p} \varepsilon_{abs} + \varepsilon_{rel} \max \left\{ \|\mathbf{F}\ell^k\|_2, \|\mathbf{s}^k\|_2, \|\mathbf{b} + \mathbf{c}\|_2 \right\} \\ \varepsilon_{dual}^k &= \sqrt{n} \varepsilon_{abs} + \varepsilon_{rel} \|\mathbf{F}\ell_s^k\|_2, \end{aligned}$$

where $p \triangleq I(I-1)$ and $n \triangleq 3$. The SAG algorithm which jointly designs the passive and active precoders, and obtains the AIRS' position is summarized in Algorithm 4, where the subscript t is used to denote the iteration index for the SCA and superscript k (again) denotes the ADMM index. The algorithm will terminate when $\|\ell_t - \ell_{t-1}\|_2 \leq \varepsilon_{SCA}$.

Algorithm 4: SCA-ADMM-GBD (SAG) Algorithm

```

1 Initialize  $t = 0$ ,  $\rho = 1$ ,  $\mathbf{b}_q$ ,  $\mathbf{u}_i$ ,  $\ell = \ell_0$ ,  $\widehat{\mathbf{h}}_{R,i}$ ,  $\widehat{\mathbf{H}}_{q,R}$ ; while
   $\|\ell_t - \ell_{t-1}\| > \varepsilon_{SCA}$  do
2   Solve  $\Theta$  and  $\mathbf{f}_{q,i}$  with  $\ell_t$  using the proposed
   GBD-based method outlined in Section III;
3   Use  $\Theta$ ,  $\mathbf{f}_{q,i}$  and  $\ell_t$  to approximate  $f_{c,q,i,j}(\ell)$ ;
4   Initialize  $\mathbf{a}^0$ ,  $\mathbf{s}^0$  and  $\ell^0 = \ell_t$ ;
5    $k = 0$ ;  $\rho^k = 1$ ;
6   while  $(\|\mathbf{r}^k\|_2 > \varepsilon_{pri}^k)$  and  $(\|\mathbf{d}^k\|_2 > \varepsilon_{dual}^k)$  do
7      $\ell$ -step using (24a);
8     B-step using (24b);
9     s-step using (24c);
10     $\lambda_s$ -step using (24d);
11     $\rho^k$  is updated according to (25);
12     $k = k + 1$ ;
13  end
14   $\ell_{t+1} = \ell^k$ ;
15   $t = t + 1$ ;
16 end

```

TABLE 1. Transmitted data size for Algorithm 1 ($L = 1$).

Line	Data	Size	Direction
3	$\mathbf{H}_{q,R}$	$N_T \times ML \times Q$	Each AP to CAP to APs
3	$\mathbf{h}_{R,i}$	$1 \times ML \times I$	Each AP to CAP to APs
5	θ	$ML \times 1$	CAP to APs
6	$\mathbf{W}_{q,i}$	$N_T \times N_T \times I$	APs to CAP
6	\mathbf{u}^k	$I(I - 1) \times 1$	APs to CAP

VII. SIMULATION RESULTS

A. COMMUNICATION OVERHEAD AND COMPUTATIONAL COMPLEXITY OF THE GBD METHOD

Table 1 and 2 show the communication overhead for Algorithms 1 and 2 (GBD-MISDP only). Note that the GBD-ILR AP-UE pairing scheme requires an exchange of information between APs. The exchanged information includes the current pairing and the ranking of UEs for each AP, which is only of size $I \times 1$ per AP.

The GBD-ILR method solves the SP in (5) multiple times and solves the SP roughly $\log_2(I/(2I/Q)) + 5 = \log_2(Q) + 4$ times per iteration with $Q \geq 4$ and large I . It is, however, still a great improvement over the GBD-MISDP as it incurs a worst-case computational complexity of I^Q per iteration, which equals the number of possible solutions for $v_{q,i}$.

B. CHANNEL MODEL FOR GBD-BASED JOINT BEAMFORMER DESIGN

It is assumed the AP-IRS channels have line-of-sight (LoS) components, while the IRS-UE channels only have a small chance of containing its LoS components, while the direct path channels between APs and UEs are blocked. An indoor factory (InF) scenario as defined in [34] is considered, with

TABLE 2. Transmitted data size for Algorithm 2 ($L = 1$).

Line	Data	Size	Direction
3	$\mathbf{H}_{q,R}$	$N_T \times ML \times Q$	Each AP to CAP
3	$\mathbf{h}_{R,i}$	$1 \times ML \times I$	Each AP to CAP
16	$\mathbf{f}_{q,i}$	$N_T \times I$	CAP to APs

the assumptions that APs and IRSs are above the clutter height and UEs are below the clutter height, which result in the AP-IRS and IRS-UE channels corresponding to the InF-HH and the InF-SH scenario, respectively. A ray-tracing model is employed. The channel provided by each cluster is given by

$$\mathbf{H}_n = \sum_{j=1}^{N_s} \sum_{m \in \mathcal{R}_j} \sqrt{\frac{P_n}{M_{r,n}}} \exp(-j\Omega_{i,j,m}) \mathbf{a}_R(\phi_{i,j,m}, \theta_{n,j,m}) \mathbf{a}_T(\phi_{n,j,m}, \theta_{n,j,m})^H,$$

where n, j and m are the cluster, sub-cluster and ray index, respectively. N_s is the number of the sub-cluster, and \mathcal{R}_j denotes the set of rays belonging to each sub-cluster. P_n denotes the cluster power, $M_{r,n}$ is the total number of rays of the n th cluster. $\Omega_{i,j,m}$ is the random initial phase. \mathbf{a}_{Rx} and \mathbf{a}_{Tx} denote the receiving and transmit spatial response vectors, where ϕ and θ are the azimuth and zenith angles. For LOS channels, the first cluster is the LOS component. All AP-IRS channels are LoS channels, and the probability of a IRS-UE channel being a LoS channel is given in [34, Table 7.4.2-1].

According to [34, Table 7.4.1-1], the path loss of the LoS and NLoS components in dB, denoted as PL_{LoS} and PL_{NLoS} , are given by

$$\begin{aligned}
 PL_{LoS} &= 31.84 + 21.5 \log_{10}(d_{3d}) + 19 \log_{10}(f_c), \\
 PL'_{NLoS} &= 32.4 + 23 \log_{10}(d_{3d}) + 20 \log_{10}(f_c), \\
 PL_{NLoS} &= \begin{cases} PL_{LoS}, & \text{for AP-IRS channels,} \\ \max(PL_{LoS}, PL'_{NLoS}), & \text{for IRS-UE channels,} \end{cases}
 \end{aligned}$$

where f_c is the carrier frequency, and d_{3d} is the distance between the transmitter and the receiver in a 3D coordinate.

Let $\mathbf{H}_n^{(q,\ell)}$ and $\mathbf{h}_n^{(\ell,i)}$ denote the channel component provided by the n th cluster, respectively, of $\mathbf{H}_{q,\ell}$ and $\mathbf{h}_{\ell,i}$. Assuming the path loss is the same for all transmitting/receiving antennas, $\mathbf{H}_{q,\ell}$ and $\mathbf{h}_{\ell,i}$ are given as

$$\begin{aligned}
 \mathbf{H}_{q,\ell} &= \sqrt{\frac{\kappa}{\kappa+1}} \mathbf{H}_1^{(q,\ell)} \mathbf{A}_{LoS}^{(q,\ell)} + \sqrt{\frac{1}{\kappa+1}} \left(\sum_{n=2}^{N_c} \mathbf{H}_n^{(q,\ell)} \right) \mathbf{A}_{NLoS}^{(q,\ell)} \\
 \mathbf{h}_{\ell,i} &= \begin{cases} \mathbf{A}_{NLoS}^{(\ell,i)} \sum_{n=1}^{N_c} \mathbf{h}_n^{(\ell,i)}, & \text{NLoS case,} \\ \sqrt{\frac{\kappa}{\kappa+1}} \mathbf{A}_{LoS}^{(\ell,i)} \mathbf{h}_1^{(\ell,i)} \\ + \sqrt{\frac{1}{\kappa+1}} \mathbf{A}_{NLoS}^{(\ell,i)} \left(\sum_{n=2}^{N_c} \mathbf{h}_n^{(\ell,i)} \right), & \text{LoS case,} \end{cases} \quad (26)
 \end{aligned}$$

where N_c denotes the number of clusters and

$$\mathbf{s}_{LoS}^{(q,\ell)} \sim \mathcal{N}(0, \xi_{LoS} \mathbf{I}_{N_T}), \in \mathbb{R}^{N_T},$$

$$\begin{aligned}
 \mathbf{s}_{NLoS}^{(q,\ell)} &\sim \mathcal{N}\left(0, \xi_{NLoS}^{(q,\ell)} \mathbf{I}_{N_T}\right), \in \mathbb{R}^{N_T}, \\
 s_{LoS}^{(\ell,i)} &\sim \mathcal{N}\left(0, \xi_{LoS}\right), \\
 s_{NLoS}^{(\ell,i)} &\sim \mathcal{N}\left(0, \xi_{NLoS}^{(\ell,i)}\right), \\
 \mathbf{A}_{LoS}^{(q,\ell)} &= 10^{\left(\left(-PL_{LoS}^{(q,\ell)} + G_{AP}\right) \mathbf{I}_{N_T} + \text{Diag}\left(s_{LoS}^{(q,\ell)}\right)\right)/20} \in \mathbb{R}^{N_T \times N_T}, \\
 \mathbf{A}_{NLoS}^{(q,\ell)} &= 10^{\left(\left(-PL_{NLoS}^{(q,\ell)} + G_{AP}\right) \mathbf{I}_{N_T} + \text{Diag}\left(s_{NLoS}^{(q,\ell)}\right)\right)/20} \in \mathbb{R}^{N_T \times N_T}, \\
 A_{LoS}^{(\ell,i)} &= 10^{\left(-PL_{LoS}^{(\ell,i)} + G_{IRS} + s_{LoS}^{(\ell,i)}\right)/20}, \\
 A_{NLoS}^{(\ell,i)} &= 10^{\left(-PL_{NLoS}^{(\ell,i)} + G_{IRS} + s_{NLoS}^{(\ell,i)}\right)/20}.
 \end{aligned}$$

$10^{\lfloor s_{LoS} \rfloor_i / 10}$ is the shadow fading in the LoS component (similar definition is given for $10^{\lfloor s_{NLoS} \rfloor_i / 10}$) that is independent of the path loss $d_{3d}^{-\alpha}$, where α denotes the path loss, which in the simulations, equals to either 2.15 or 2.3. Assuming all antenna array configurations at the APs are uniform linear arrays (ULAs) and those for the IRSs are uniform planar arrays (UPAs). Then their spatial response vectors can be expressed as

$$\begin{aligned}
 \mathbf{a}_{ULA}(\phi) &= \frac{1}{\sqrt{(N_x)}} \left[1 \quad \cdots \quad \exp(-j2\pi(N_x - 1)\Delta_x \sin(\phi)) \right]^T \\
 \mathbf{a}_{UPA}(\phi, \theta) &= \frac{1}{\sqrt{(N_x N_y)}} \left[1 \quad \cdots \quad \exp(-j2\pi(N_x - 1)\Delta_x \sin(\theta)) \right. \\
 &\quad \cos(\phi) \exp(-j2\pi \Delta_y \sin(\theta) \sin(\phi)) \quad \cdots \\
 &\quad \exp(-j2\pi(N_x - 1)\Delta_x \sin(\theta) \cos(\phi)) \\
 &\quad \quad -j2\pi \Delta_y \sin(\theta) \sin(\phi) \quad \cdots \\
 &\quad \exp(-j2\pi(N_y - 1)\Delta_y \sin(\theta) \sin(\phi)) \quad \cdots \\
 &\quad \left. \exp(-j2\pi(N_x - 1)\Delta_x \sin(\theta) \cos(\phi)) \right. \\
 &\quad \quad \left. -j2\pi(\Delta_y - 1)\Delta_y \sin(\theta) \sin(\phi) \right]^T,
 \end{aligned}$$

where \mathbf{a}_T equals to either $\mathbf{a}_{ULA} \in \mathbb{C}^{N_x}$ or $\mathbf{a}_{UPA} \in \mathbb{C}^{N_x \times N_y}$ and \mathbf{a}_R equals to \mathbf{a}_{UPA} for the downlink channel between the APs and the IRSs. N_x and N_y denote the number of antennas along the x or y -axis, respectively. Δ_x and Δ_y denote the distance between antennas along the x or y -axis in the unit of wavelength of the signal. All the parameters not explained above are generated according to [34, Sec. 7.5].

C. SIMULATION SETUP

The AP and IRS panels are set vertically to the ground. It is assumed each IRS panel can only reflect signals with AoD between $[-90^\circ, 90^\circ]$, i.e., IRSs can only reflect signals from the direction it is facing. Placement of the APs, IRSs, and UEs are described by 2D coordinates with parameters d_y, d_1, d_2, d_3 , as shown in Fig. 1. The APs are uniformly placed on a line segment $(0, 0)$ to $(0, d_y)$. IRSs can have an orientation of either 90° or -90° and a pair of IRSs with different orientations are grouped as a set. IRS sets are uniformly placed along the line segment $(d_1, 0)$ to (d_1, d_y) , with the exception at both ends, where only one IRS panel is placed. UEs are randomly placed in a rectangle with corners $(d_1 + d_2, 0)$, $(d_1 + d_2, d_y)$, $(d_1 + d_2 + d_3, 0)$, and $(d_1 +$

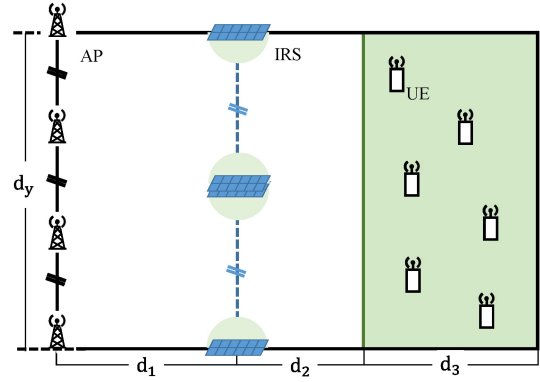


FIGURE 1. An illustration of the simulation setup with 4 APs and 4 IRSs. UEs are randomly placed in the green area and the half-circles indicate the orientation of the IRSs.

$d_2 + d_3, d_y)$. Due to the orientation of the IRSs, each AP and UE has access to $M(L/2 + 1)$ IRS reflecting elements, which is roughly half of the number of total IRS reflecting elements in the system. The set up is illustrated in Fig. 1. All parameters are summarized in Table 3. I_{th} is adjusted according to other parameters such that the received signal power and the interference power would be of the same order achieving a signal-to-interference power ratio (SIR) of -3dB to 25dB as indicated in [39]. P^q is chosen such that the received signal power at each user would be above the reference sensitivity power level in [24]. The simulation is implemented in MATLAB. CVX [41], [42] is used to solve (5) and (6). The branch-and-bound (BNB) algorithm of YALMIP [43] is used to solve (11). The MOSEK [44] solver is used in both toolboxes. The simulation is done on a machine with Intel Core i9-11900K CPU at 3.5GHz and 32GB RAM.

D. CONVERGENCE BEHAVIOR OF THE GBD METHOD

Fig. 2 shows the convergence behavior of Algorithm 1 with $Q = 6, I = 16, L = 2$. Fig. 2(a) shows the objective value produced by (5) (labeled as LB) and (8) (labeled as UB) at each iteration and Fig. 2(b) is the sum rate at each iteration. Both results are averaged over 100 experiments. It can be seen in the results that the algorithm converges between 2 to 5 iterations. The run time, though, does increase with the problem size, i.e., when N_T, I, Q, M or L increases. It can also be observed that the convergence in the objective function does translate to the maximization of the sum rate.

E. JOINT AP-IRS BEAMFORMING DESIGN WITH AP-UE PAIRING

Figs. 3 shows two sets of sum rate vs. number of UEs results; one when Q is fixed and the other when L is fixed. 100 experiments were conducted per data point. Increasing Q or L both lead to growth in the sum rate, but the latter is much more effective than the former as the performance gap between different values of L is greater than that between different values of Q . This shows the effectiveness of IRS.

TABLE 3. Simulation parameters for the GBD-based algorithm.

Parameter	value	unit
N_T	8	(No unit)
M	4×4	(No unit)
κ	$\sim \mathcal{N}(7, 8)$	dB
ξ_{LoS}	4	dB
$\xi_{NLoS}^{(q,\ell)}, \forall q \in \mathcal{Q}, \forall \ell \in \mathcal{L}$	4	dB
$\xi_{NLoS}^{(\ell,i)}, \forall \ell \in \mathcal{L}, \forall i \in \mathcal{I}$	5.9	dB
I_{th}	10^{-8}	W
Noise power	-160	dBW/Hz
$P^q, \forall q \in \mathcal{Q}$	80	dBm
Carrier frequency f_c	2.5	GHz
Antenna gain of AP G_{AP}	8	dBi
Antenna gain of IRS G_{IRS}	5	dBi
$\Delta_{AP}, \Delta_{IRS,x}, \Delta_{IRS,y}$	0.5	λ
Height of AP	10	m
Height of UE	2	m
Height of IRS	10	m
Effective Clutter Height	5	m
Clutter Density	0.3	(no unit)
Ceiling Height	15	m
d_y	50	m
d_1	15	m
d_2	90	m
d_3	15	m
Convergence threshold ε (Algorithms 1 and 2)	-80	dBm
γ (Algorithm 3)	1.5	(No unit)
Number of Monte Carlo Simulations	100	(No unit)

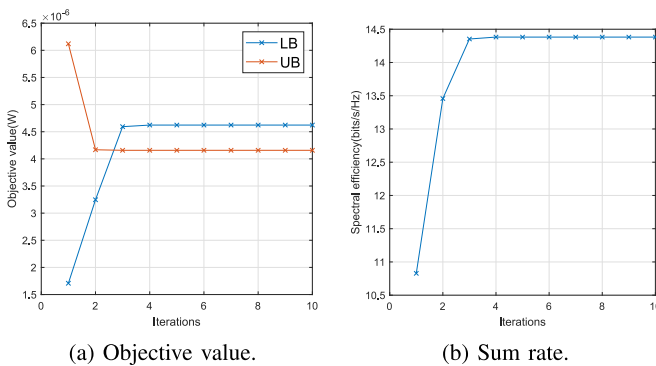


FIGURE 2. Convergence behavior of Algorithm 1 ($Q = 6, I = 12, L = 2$).

Increasing the number of APs increases the received signal power per user as more transmit power is available for usage in the whole system. This, however, also translates to more leakage interference per user, but still result in a net gain in terms of spectrum efficiency.

Fig. 4 shows the sum rate vs. number of UEs performance for a system with $Q = 8$ and $L = \{1, 2\}$. In the figure, the performance of the following four algorithms are compared:

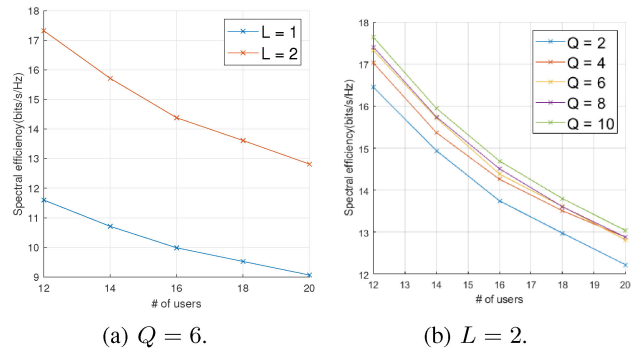


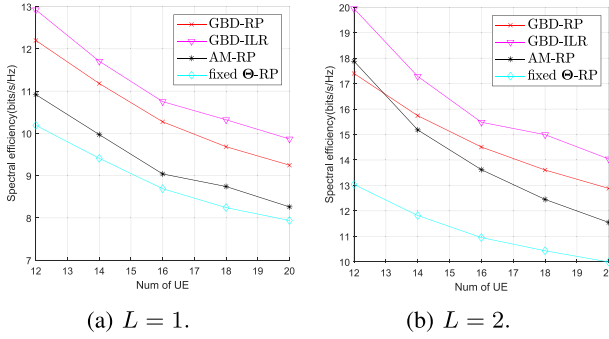
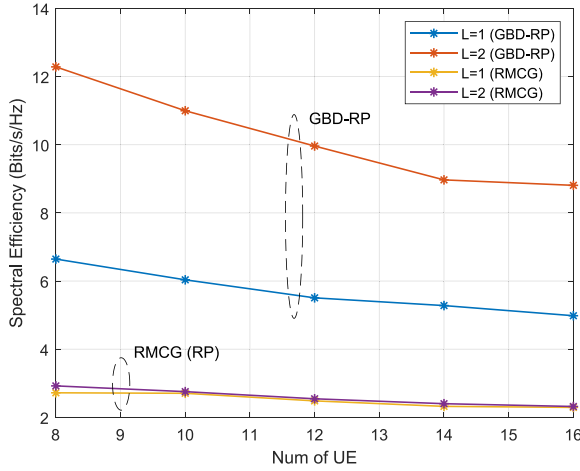
FIGURE 3. Sum rate comparison for different number of APs and IRS using GBD-RP.

- **GBD-ILR algorithm:** Uses Algorithm 2, that calls Algorithm 3 to solve the joint beamformer and AP-UE pairing problems.
- **GBD-RP:** Uses Algorithm 1 to solve the joint beamformer design problem with random AP-UE pairing \mathcal{P} , where UEs are randomly assigned to APs, with each AP serving the same number of UEs.
- **AM-RP:** Alternating maximization (AM) with random pairing algorithm solves (1) by alternatively updating one variable while fixing the other. The solutions $\mathbf{f}_{q,i}^*$ and Θ^* are obtained by choosing the best objective value (i.e., the received signal power) after running the AM algorithm for 20 iterations as the algorithm does not always converge for this problem.
- **Fixed Θ -RP:** Fixing $\Theta = \mathbf{I}_{ML}$ and designs $\mathbf{f}_{q,i}$ by solving (5) with randomization with random AP-UE pairing.
- **RMCG:** This is the Riemannian manifold conjugate gradient method that was proposed in [16] to obtain θ , that guarantees at least a local optimal solution. The active beamformer $\mathbf{f}_{q,i}$ is solved using the method stated in [16].

According to Figs. 4(a) and 4(b), the proposed GBD-ILR joint beamformer design always performs the best. The GBD-ILR method converges between 2 to 5 iterations, which displays similar convergence behavior to the GBD-RP method. It can be observed from Fig. 4 that optimal AP-UE pairing does have an advantage over random pairing, and the advantage grows larger as L increases. In the case of $L = 2$, even though AM-RP can outperform GBD-RP when the number of users is less than 13, it often fails to converge as the system size grows, i.e., when N_T , Q , or I are large, which indicates its lack of ability to scale.

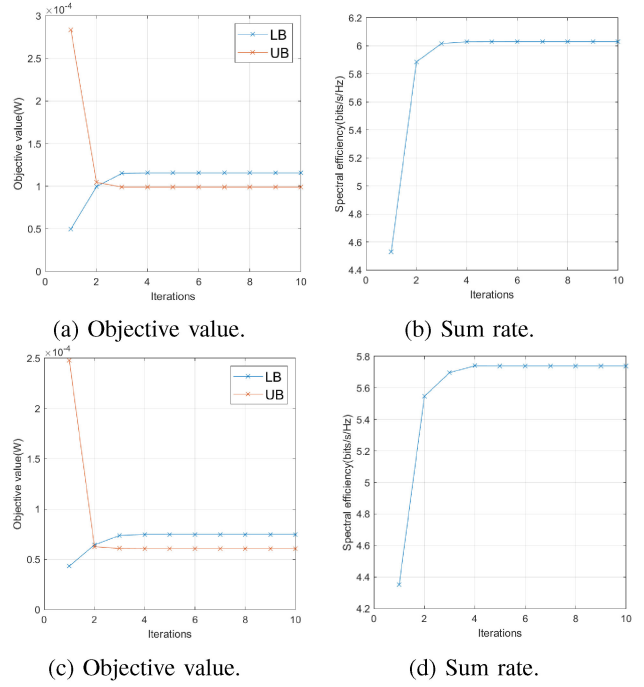
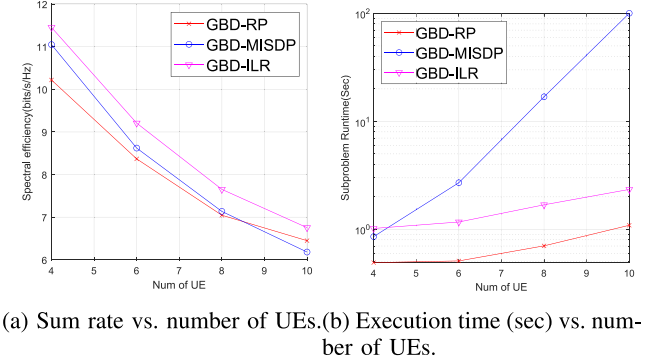
The proposed GBD-RP algorithm is compared to the method in [16] in terms of sum rate in Fig. 5 for $L = 1$ and 2, which only deals with a single AP system, with multiple IRSs and users. As seen in the figure, the proposed GBD-RP method always outperforms the method in [16] despite the number of users or IRSs.

Fig. 6 depicts the convergence behavior of Algorithm 2 in terms of received signal power (objective value) and spectral efficiency. It is clear that both GBD-MISDP and


 FIGURE 4. GBD, GBD-ILR, AM, and fixed-IRS system ($\Theta = \mathbf{I}_{ML}$) comparison ($Q = 8$).

 FIGURE 5. Sum rate vs. number of UEs performance for GBD-RP and the method in [16] with different number of IRSs. (a) $L = 1$. (b) $L = 2$.

GBD-ILR converge in about 2-5 iterations, which is similar to the convergence behavior of Algorithm 1 shown in Fig. 2. Moreover, GBD-MISDP converges to a higher objective and spectral efficiency value than the GBD-ILR. This is supported in

Fig. 7, which compares the sum rate and simulation time performance for GBD-MISDP, GBD-ILR and GBD-RP. Note that only a small size system (i.e., $Q = 2$, $\max(I) = 10$, $L = 2$, $M = 2 \times 2$) can be simulated due to the exorbitant amount of time needed for the GBD-MISDP. Due to the small system size, $I_{th} = 10^{-6}$ is used. From the figure, it is clear the GBD-ILR and GBD-RP require much less time to find the solution than the GBD-MISDP. The execution time of GBD-MISDP grows exponentially with system size, while that of GBD-ILR grows at the same speed with GBD-RP (i.e., the gap between the two curves stays the same.). This is consistent with the analysis in Section VII-A. The sum rate performance for GBD-ILR is very close to, and sometimes outperforms that of GBD-MISDP. This does not contradict that the GBD-MISDP finds the global optimal solution for (9) because the objective is the received signal power. Moreover, the GBD-ILR is able to outperform the GBD-MISDP in terms of sum rate because the former achieves lower interference power.


 FIGURE 6. Convergence behavior of Algorithm 2 ($Q = 2$, $L = 1$, $M = 2 \times 2$, $I_{th} = 10^{-6}$) using (a) GBD-MISDP in terms of objective value, (b) GBD-MISDP in terms of spectral efficiency, (c) GBD-ILR in terms of objective value, and (d) GBD-ILR in terms of spectral efficiency.

 FIGURE 7. (a) Sum rate vs. number of UEs, and (b) Execution time (sec) vs. number of UEs performance for GBD-MISDP, GBD-ILR and GBD-RP with $Q = 2$, $L = 2$, $M = 2 \times 2$, $I_{th} = 10^{-6}$.

F. RESULTS USING THE SAG ALGORITHM

In the simulation using the proposed SAG method, the centroid of the APs is 100 m away from that of the UEs in the x -axis. The channels are generated using (12), where the path loss exponent $\alpha = 2.1$, which corresponds to the UMi scenario in [34]. The mean and the variance of the entries of $\mathbf{H}_{q,R,w}$ and $\mathbf{h}_{R,i,w}$ equals $\mu = \sqrt{\frac{\kappa}{2(\kappa+1)}}$, and $\sigma = \sqrt{\frac{1}{2(\kappa+1)}}$, with $\kappa = 10$ dB, for all $q = 1, \dots, Q$ and $i = 1, \dots, I$. As shown in Fig. 8, the APs are centered around the origin and positioned on the y -axis at 5 m apart, with a height of 10 m. The UEs are uniformly distributed in a rectangular box of size 15×20 m, centered at 100 m, with a height of 1.5 m. The position of the AIRS varies according to the different experiments below. It is assumed that the only paths for

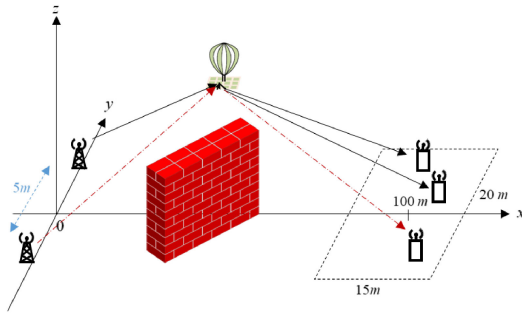


FIGURE 8. AIRS configuration consisting of 2 APs and 3 UEs. Distance between APs are 5m with a height of 10m, and are aligned on the y-axis. The UEs are uniformly distributed in a rectangular box of size 15 × 20m, centered at 100m, with a height of 1.5m. The position of the AIRS varies.

TABLE 4. Simulation parameters for AIRS.

Parameter	value	unit
N_T	8	(No unit)
M	32	(No unit)
L	1	(no unit)
Path loss exponent α	2.1	(no unit)
Variance of shadow factor $\xi_{q,n}, \xi_{i,m}$	4	dB
Line of sight factor κ	10	dB
I_{th}	10^{-7}	W
Noise power $P^q, \forall q \in \mathcal{Q}$	-160	dBW/Hz
Carrier frequency f_c	2.5	GHz
Antenna gain of AP G_{AP}	10	dBi
Antenna gain of IRS G_{IRS}	10	dBi
$\Delta_{AP}, \Delta_{IRS,x}, \Delta_{IRS,y}$	0.5	λ
Height of AP	10	m
Height of UE	1.5	m
Height of IRS	30, 60, 90	m
Distance between APs	5	m
SCA convergence threshold ε_{SCA}	0.01	(no unit)
ADMM absolute tolerance ε_{abs}	0.1	(no unit)
ADMM relative tolerance ε_{rel}	0.1	(no unit)
GAM momentum factor β	0.7	(no unit)
GAM step size γ	0.5	(No unit)
Number of Monte Carlo Simulations	100	(No unit)

transmission from the APs to the UEs are via the AIRS as other paths are blocked. Table 4 summarizes the simulation parameters for the SAG algorithm.

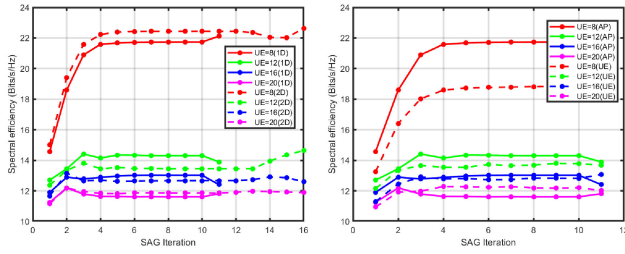
Even though the proposed algorithm can obtain the 3D coordinates of the AIRS, the SAG algorithm is only used to produce 1 and 2D coordinates of the AIRS in the simulations below. However, the spectral efficiency at different height will be shown. The AP-UE pairing is done via random pairing with each AP serving the same number of UEs.

Below are the methods that are used in the experiments below.

- **1D-SAG:** The SAG algorithm outlined in Algorithm 4 is used to obtain the 1D location of the AIRS on the x -axis and design the beamformers jointly.
- **2D-SAG:** The SAG algorithm outlined in Algorithm 4 is used to obtain the 2D location of the AIRS on the x and y -axis and design the beamformers jointly.
- **1D-Zhou (ZF) and (GBD):** The method in [27] is used to obtain the 1D location of the AIRS on the x -axis. The method can only render the 1D coordinate of the AIRS. Since [27] only considered 1 UE, its joint beamformer solution cannot be used in the multiple UEs case considered herein. Hence, for fair comparison, the method is used in conjunction with the zero-forcing (ZF) precoder and the GBD method in Section IV. Note that the ZF precoder can only mitigate interference caused by the transmission by the same AP.
- **1D- and 2D-search with (ZF) and (GBD):** These serve as an upper bound for performance. For the 1D-search, the x -axis is sampled from 0 to 107.5 m, at 1 m apart. At each grid point, a single channel realization will be generated. The one with the smallest path loss will be selected to be the location of the AIRS. 100 channel realizations will then be generated at this point. Either the ZF or GBD beamformer design will be used to compute the spectral efficiency for each channel realization and the results will be averaged. A similar method is used for the 2D-search, except the search area equals $x = 0$ to 107.5 m, and $y = -20$ to 20 m. The grid points are 0.1 m apart in both x and y direction.

Fig. 9 depicts the convergence behavior of the SAG algorithm in terms of spectral efficiency. It is clear from Fig. 9(a) that the 2D-SAG will eventually converge to a slightly higher spectral efficiency than the 1D-SAG. From Fig. 9(b), when the number of UE is small, initializing the 1D-SAG at the centroid of the APs outperforms initialization at the centroid of the UEs. When the number of UEs increases, the trend is reversed. This is also supported and further elaborated in Fig. 11 when the sum rate vs. number of UEs performances are compared at different initialization points.

It can be observed in Fig. 10 that the sum rate of 1D-SAG and 2D-SAG are tightly bounded by the 1D-search (GBD) and 2D-search (GBD) methods, respectively, which implies the proposed approaches can indeed achieve optimal performance. In addition, the proposed SAG method performs the best among all the methods compared. It can also be seen that the 1D-Zhou (GBD) method outperforms the 1D-Zhou (ZF) method, thus, ensuring the proposed GBD design is better than the ZF method. The 1D-search (ZF) and 2D-search (ZF) is able to upper bound the 1D-Zhou (ZF) method but their performances do not match their GBD counterparts due to existing interference caused from



(a) Convergence behavior of 1D- and 2D-SAG. (b) Convergence behavior of 1D-SAG with different initialization.

FIGURE 9. Convergence behavior of Algorithm 4 in terms of sum rate. Number of APs equals to 4. Height of the AIRS equals 30 m. The centroid of the APs' location is used to initialize the SAG algorithm in (a). The centroid of the APs and UEs are used to initialize the 1D-SAG algorithm in (b). 100 random channel and random UE location realizations were used.

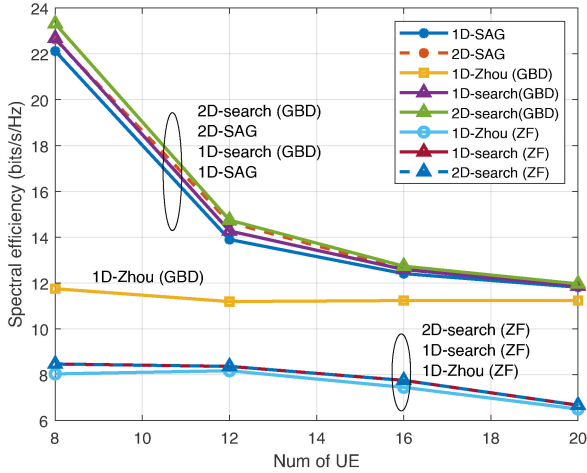


FIGURE 10. Sum rate vs. number of UEs comparison for 1D-SAG, 2D-SAG, 1D-Zhou (GBD), 1D-Zhou (ZF), 1D-search (ZF), and 2D-search (ZF). Number of APs equals to 4. Height of the AIRS equals 30 m. The centroid of the APs' location is used to initialize the SAG algorithm. 100 random channel and random UE location realizations were used.

transmissions from the q th AP to those from the p th AP, for $q \neq p$.

Upon analyzing the cost function in (15), it is found that there are many peaks near the locations of the APs and UEs. However, in the considered scenario where an AP can serve multiple UEs, the peaks near the APs have greater amplitude than those near the UEs. In addition, if the height of the AIRS is less than half of the distance between the (centroid of the) APs and the UEs, then the objective in (15) will have two peaks; one is near the APs and the other is near the UEs [26, Fig. 4], so that when the SAG algorithm is initialized at the AP (UE), the final optimal location of the AIRS will be near the AP (UE), assuming the above premise is true. Since (21) involves maximizing the objective function in (15), and due to the nonconvex nature of the problem, it is advantageous to initialize the algorithm that solves (21) to a point near the APs. Fig. 11 compares the sum rate performance of using the proposed SAG, in 1D only, initialized at two different points, one at the centroid of the APs (labeled as 1D-SAG (AP)), and the other at the centroid of the UEs (labeled as

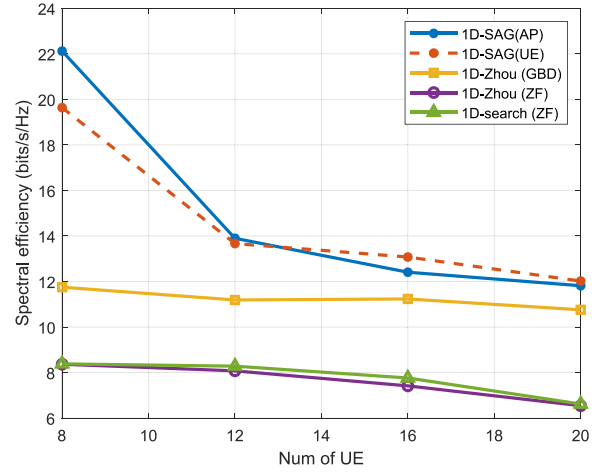


FIGURE 11. Sum rate vs. number of UEs comparison between 1D-SAG (AP) and 1D-SAG (UE). Sum rate performance for 1D-Zhou (GBD), 1D-Zhou (ZF) and 1D-search (ZF) are also included. Number of APs equals to 4. Height of the AIRS equals 30 m. 100 random channel and random UE location realizations were used.

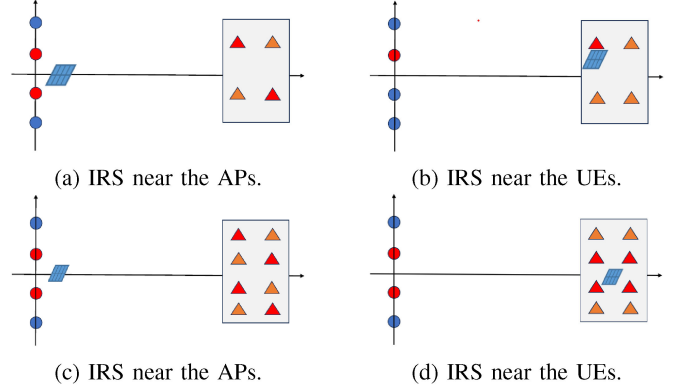


FIGURE 12. Illustration of how different initializations of the AIRS' location impact the performance as the number of UEs increases. (a) and (b) are associated with the case with fewer UEs. (c) and (d) are associated with the case with more UEs. Circles and triangles represent AP and UE, respectively. The (red triangle) UEs are served by the (red circle) APs, which has less path loss in the AP-AIRS links, than those of the (blue circle) APs, which serve the (orange triangle) UEs. The AIRS is located at its optimal location.

1D-SAG (UE)). Comparison is also carried out with 1D-Zhou (ZF) and (GBD), and 1D-search (ZF). From the figure, when $I = 8$, 1D-SAG (AP) outperforms 1D-SAG (UE) by about 2 bps/Hz. However, at and after $I = 12$, their sum rate performances are similar. This can be explained by Fig. 12. Suppose there are 4 APs in the system. Given the fact that random pairing is used, and each AP must serve the same (or similar) number of UEs, then in Figs. 12(a) and 12(b), which have less users than in Figs. 12(c) and 12(d), 1 AP will serve 1 user. Based on the configuration under consideration, there will be 2 APs (denoted as red circles) that are much closer to the (final optimal location of the) AIRS than the other two (denoted as blue circles) if the AIRS location is initialized at the centroid of the APs, given that the height of the AIRS is 30 m, which is less than half of the distance between the centroids of the APs and UE. Since the path losses between the (red circle) APs and the AIRS are much less than those

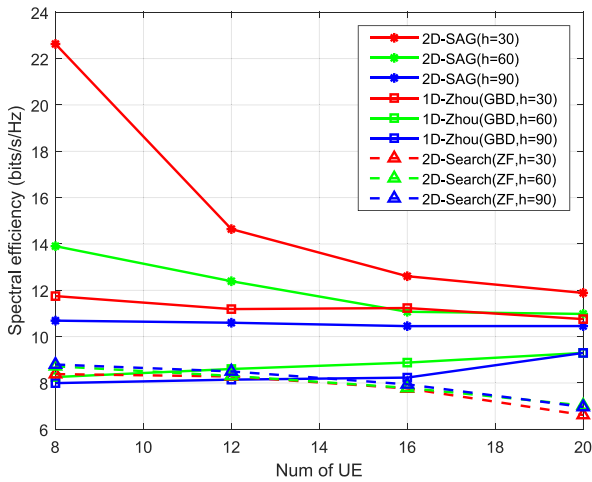


FIGURE 13. Sum rate vs. number of UEs comparison for 2D-SAG, 1D-Zhou (GBD) and 2D-search (ZF), when the height of the AIRS equals 30, 60 and 90 m. Number of APs equals to 4. The centroid of the APs' location is used to initialize the SAG algorithm. 100 random channel and random UE location realizations were used.

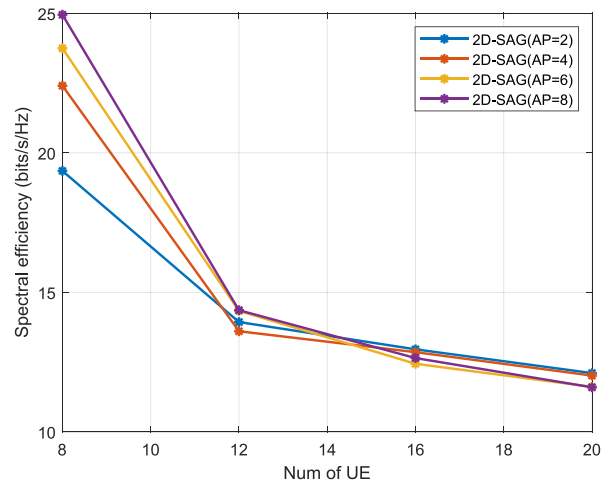


FIGURE 14. Sum rate vs. number of UEs for 2D-SAG with different number of APs. The centroid of the APs' location is used to initialize the SAG algorithm. Height of the AIRS equals 30 m. 100 random channel and random UE location realizations were used.

of the (blue circle) APs, the spectral efficiency of the UEs (red triangle) served by these (red circle) APs will be much higher than the other two (orange triangle) UEs, which are served by the (blue circle) APs. Consider the case where the initialization of the AIRS location is at the centroid of the UEs as shown in Fig 12(b), since there are fewer UEs than in Figs. 12(c) and 12(d), it is likely that the AIRS will be close to only one UE. Hence, when the number of the UEs is small, 1D-SAG (AP) is able to outperform 1D-SAG (UE). On the other hand, when the number of UEs increases as shown in Figs. 12(c) and 12(d), when the initialization is at the centroid of the APs (Fig. 12(c)), the two (red circle) APs that are closer to the AIRS than the two (blue circle) APs will serve 4 (red triangle) UEs, which will contribute to higher spectral efficiency than the 4 (orange triangle) UEs that are served by the 2 (blue circle) APs. When the AIRS location is initialized at the centroid of the UEs (Fig. 12(d)), there will be 4 (red triangle) UEs that will have higher spectral efficiency than the other 4 (orange triangle) UEs since they are closer to the AIRS. Nevertheless, since there are 4 (red circle) UEs that will contribute to the majority of the spectral efficiency in Figs. 12(c) and 12(d), hence, the sum rate for the different initializations will be the same.

Fig. 13 shows that the performance of 2D-SAG, 1D-Zhou (GBD) and 2D-search (ZF) when the AIRS is placed at different height. In general, when the AIRS is closest to the APs (at 30m), the performance is optimal, except for 2D-search (ZF). This is because the ZF precoder cannot cancel the interference caused by different APs. Also, the performance gap at different height decreases as the number of UEs increases because of the increase in interference. The performance of the 2D-SAG when the AIRS' height equals to 30 m is the best amongst all methods because the APs-AIRS channels have the lowest path loss compared to other heights.

Fig. 14 shows the sum rate vs. number of UEs performance of 2D-SAG with different number of APs. At $I = 8$, the performance improves monotonically as the number of APs increases. However, at $I = 12$ and 16, the performance gap between different number of APs decreases gradually. Finally, at $I = 20$, it can be observed that the performance of fewer AP outperforms that of more APs. This can be explained in Fig. 15 which shows the final optimal location of the AIRS, which is always close to the APs because the initial location of the AIRS is at the centroid of the APs, with different number of APs and UEs. When the number of the APs and UEs are small (Fig. 15(a)), each AP will serve 2 UEs. As the number of APs grows as shown in Fig. 15(b), each AP only needs to serve one UE. Even though the (blue circle) APs will contribute less to the sum rate than the (red circle) APs, the lesser load on each AP, compared to the case in Fig. 15(a), will lead to a system that has higher sum rate. On the other hand, when the number of UEs increases to 8 as shown in Figs. 15(c) and 15(d), when the number of AP is small (Fig. 15(c)), each AP will need to serve 4 UEs. Even when the number of APs increases to 4 (Fig. 15(d)), so that each of them only needs to serve 2 UEs, with 4 of these (orange triangle) UEs being served by the (blue circle) APs, the lesser loading on the APs is enough to compensate the reduced spectral efficiency from the (orange triangle) UEs. Hence, the sum rate performance of the AIRS-assisted system with increased number of APs is similar to that of lesser APs.

Fig. 16 shows the convergence curve for the ℓ -step in the first iteration (i.e., t and k both equal 0) of the SAG algorithm measured by $\|\ell^i - \ell^{i-1}\|_2$ vs. gradient ascent or GAM iteration i . It is clear from Fig. 16(a) that gradient ascent ($\beta = 0$) does not converge when ℓ^0 is initialized at the centroid of the APs while the GAM method ($\beta = 0.7$) does. This is because the norm of gradient of the objective

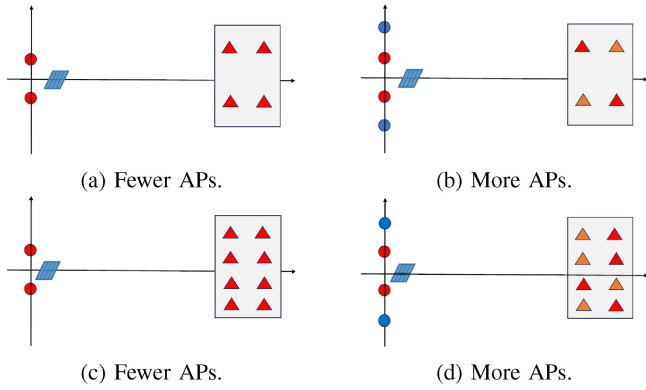


FIGURE 15. Illustration of how the number of APs impact the performance as the number of UEs increases. The representation as same as Fig. 12.

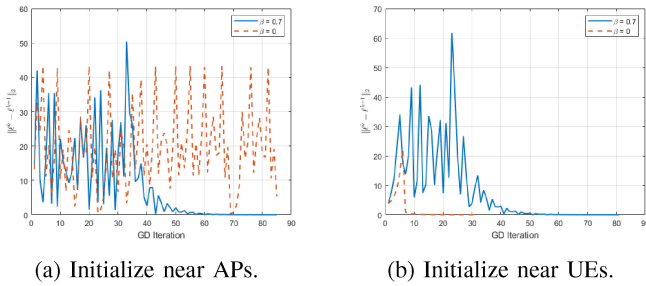


FIGURE 16. Convergence behavior of the l -step in 1D-SAG in the first SAG iteration measured as $\|l^l - l^{l-1}\|_2$, where l denotes the gradient ascent or GAM index. $Q = 4$, $l = 8$. $\beta = 0$ corresponds to the gradient ascent. $\beta = 0.7$ corresponds to the GAM, with β denoting the weighting factor for the momentum term. The curves are the result of running only one experiment. (a) The initial l_0 is at the centroid of the APs. (b) The initial l_0 is at the centroid of the UEs.

in (24a) in this case is too large, preventing the gradient ascent algorithm in reaching its stationary point. The GAM, on the other hand, can compensate due to the existence of the momentum. This also explains why both methods converged when l^0 is initialized at the centroid of the UEs because the norm of the gradient is smaller than the first case. To understand why the gradient is larger when l^0 is initialized near the APs (and remains there after convergence) compared to the case when l^0 is initialized near the UEs (and remains there after convergence), assume there is a single AP and 3 UEs. Then the objective from (15) is

$$f_0(l) \propto \frac{1}{\|b_1 - l\|_2 \|u_1 - l\|_2} + \frac{1}{\|b_1 - l\|_2 \|u_2 - l\|_2} + \frac{1}{\|b_1 - l\|_2 \|u_3 - l\|_2}.$$

Since there are three $\|b_1 - l\|_2$ terms in common, when l^0 is initialized near the APs, $\|f_0(l)\|_2$, and in turn $\|\nabla f_0(l)\|_2$, will be large. On the other hand, when l^0 is initialized near the UEs, the final location will likely be near the UEs and close to one of the UEs. Assume user 1 is closest to the AIRS, then only $\|u_1 - l\|_2$ will be small. Therefore the norm of $f_0(l)$ will be smaller than the case when the AIRS is closer to the APs.

VIII. CONCLUSION

In the first part of this work, the joint beamforming design problem for multi-AP, multi-IRS and multi-UE MISO system is considered, with single IRS being a special case. Two problems were considered. The first considered the joint beamforming problem with predefined AP-UE pairing and the second incorporated the user selection problem. The first problem can be solved using the GBD algorithm after employing mathematical reformulation and SDR. The second problem can be formulated as a MISDP problem such that the proposed GBD-MISDP method can be used to find the global, or near global, optimum solution. In addition, a heuristic AP-UE pairing algorithms is proposed to solve the pairing problem that requires less computations than the GBD-MISDP method.

With the exception of the GBD-MISDP method, all proposed algorithms (GBD-RP and GBD-ILR) can be applied distributively and in parallel when solving for the active precoder. The amount of data overhead connected to distributed design is analyzed. Out of all proposed algorithms, GBD and GBD-MISDP have finite ε -convergence guarantee and are guaranteed to at least attain the local optimal solution due to the use of the GBD algorithm. For GBD-ILR, it is guaranteed that the UB and $UB - LB$ is nonincreasing. The GBD-RP and GBD-ILR algorithms show similar convergence behavior in the simulation and take 2 to 5 iterations to converge. As a result, the proposed methods have shown to outperform the alternating maximization approach and when the phase shifters at the IRS are fixed.

In the second part, the AIRS position problem is incorporated along with the beamformer design problem and are determined jointly using the proposed SAG algorithm. It is shown that the SAG method can obtain sum rate performance that is close to that of the exhaustive search method, thus proving its optimality. Despite having the need to solve a nonconvex problem inside the SAG, it is shown that the algorithm always converges through careful initialization and the use of the GAM.

REFERENCES

- [1] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless networks," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.
- [2] X. Yang, C.-K. Wen, and S. Jin, "MIMO detection for reconfigurable intelligent surface-assisted millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1777–1792, Aug. 2020, doi: 10.1109/JSAC.2020.3000822.
- [3] M. Di Renzo et al., "Smart radio environments empowered by AI reconfigurable meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 129, no. 1, pp. 1–20, Dec. 2019.
- [4] W. Tang et al., "MIMO transmission through reconfigurable intelligent surface: System design, analysis, and implementation," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2683–2699, Nov. 2020.
- [5] J. Lyu and R. Zhang, "Hybrid active/passive wireless network aided by intelligent reflecting surface: System modeling and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7196–7212, Nov. 2021.
- [6] J. Lyu and R. Zhang, "Spatial throughput characterization for intelligent reflecting surface aided multiuser system," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 834–838, Jun. 2020.

- [7] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
- [8] X. Ma, S. Guo, H. Zhang, Y. Fang, and D. Yuan, "Joint beamforming and r Equationqref multi-user communication systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 3269–3283, May 2021.
- [9] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for intelligent reflecting surface enhanced wireless networks," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, 2019, pp. 1–6.
- [10] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H.V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, Aug. 2020.
- [11] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82599–82612, 2019.
- [12] C. Huang, A. Zappone, M. Debbah, and C. Yuen, "Achievable rate maximization by passive intelligent mirrors," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2018, pp. 3714–3718.
- [13] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.
- [14] K. Shen and W. Yu, "Fractional programming for Communication systems part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018, doi: [10.1109/TSP.2018.2812733](https://doi.org/10.1109/TSP.2018.2812733).
- [15] K. Shen and W. Yu, "Fractional programming for communication systems part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, May 2018, doi: [10.1109/TSP.2018.2812748](https://doi.org/10.1109/TSP.2018.2812748).
- [16] Z. Li, M. Hua, Q. Wang, and Q. Song, "Weighted sum-rate maximization for multi-IRS aided cooperative transmission," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1620–1624, Oct. 2020.
- [17] J. Wang, H. Wang, Y. Han, S. Jin, and X. Li, "Joint Transmit beamforming and phase shift design for reconfigurable intelligent surface assisted MIMO systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 354–368, Jun. 2021.
- [18] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44304–44321, 2021.
- [19] J. Lin, Y. Zout, X. Dong, S. Gong, D. T. Hoang, and D. Niyato, "Deep reinforcement learning for robust beamforming in IRS-assisted wireless communications," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.
- [20] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839–1850, Aug. 2020.
- [21] S. Huang, Y. Ye, M. Xiao, H. V. Poor, and M. Skoglund, "Decentralized beamforming design for intelligent reflecting surface-enhanced cell-free networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 673–677, Mar. 2021.
- [22] V. Kumar, Z. Ding, and M. F. Flanagan, "On the performance of downlink NOMA in underlay spectrum sharing," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4523–4540, May 2021.
- [23] Y. Zhang et al., "Beyond cell-free MIMO: Energy efficient reconfigurable intelligent surface aided cell-free MIMO communications," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 412–426, Jun. 2021.
- [24] "5G, NR, user equipment (UE) radio transmission and reception; Part 1: Range 1 standalone; (Release 16), version 16.5.0." 3GPP, Sophia Antipolis, France, Rep. TS 38.101-1, 2020.
- [25] S. Zeng, H. Zhang, B. Di, Z. Han, and L. Song, "Reconfigurable intelligent surface (RIS) assisted wireless coverage extension: RIS orientation and location optimization," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 269–273, Jan. 2021.
- [26] H. Lu, Y. Zeng, S. Jin, and R. Zhang, "Enabling panoramic full-angle reflection via aerial intelligent reflecting surface," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2020, pp. 1–6.
- [27] T. Zhou, K. Xu, X. Xia, W. Xie, and J. Xu, "Achievable rate optimization for aerial intelligent reflecting surface-aided cell-free massive MIMO system," *IEEE Access*, vol. 9, pp. 3828–3837, 2021.
- [28] Y. Su, X. Pang, S. Chen, X. Jiang, N. Zhao, and F. R. Yu, "Spectrum and energy efficiency optimization in IRS-assisted UAV networks," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6489–6502, Oct. 2022.
- [29] X. Song, Y. Zhao, Z. Wu, Z. Yang, and J. Tang, "Joint trajectory and communication design for IRS-assisted UAV networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1538–1542, Jul. 2022.
- [30] Y. Cai, Z. Wei, S. Hu, D. W. K. Ng, and J. Yuan, "Resource allocation for power-efficient IRS-assisted UAV communications," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Dublin, Ireland, 2020, pp. 1–7.
- [31] S. Jiao, F. Fang, X. Zhou, and H. Zhang, "Joint beamforming and phase shift design in downlink UAV networks with IRS-assisted NOMA," *J. Commun. Inf. Netw.*, vol. 5, no. 2, pp. 138–149, Jun. 2020.
- [32] K. Li, K. Zhao, M. F. Khan, P.-H. Ho, and L. Peng, "UAV-mounted intelligent reflecting surface (IRS) MISO communications," in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, Urumqi, China, 2022, pp. 62–66.
- [33] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.
- [34] "5G Study on channel model for frequencies from 0.5 to 100 GHz; (Release 16), Version 16.1.0." 3GPP, Sophia Antipolis, France, Rep. TR 38.901, 2020.
- [35] M. Servetnyk and C. C. Fung, "Precoding and selection for coordinated multipoint transmission in fronthaul-constrained cloud-RAN," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 51–55, Jan. 2020.
- [36] "Study on scenarios and requirements for next generation access technologies; Version 16.0.0." 3GPP, Sophia Antipolis, France, Rep. TR 38.913, 2020.
- [37] A. M. Geoffrion, "Generalized benders decomposition," *J. Optim. Theory Appl.*, vol. 10, pp. 237–260, Oct. 1972.
- [38] J. Zhao, T. Q. S. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2762–2775, Jun. 2013, doi: [10.1109/TWC.2013.050613.120825](https://doi.org/10.1109/TWC.2013.050613.120825).
- [39] "E-mail discussion summary of the large scale calibration," 3GPP, Sophia Antipolis, France, Rep. R1-165974.2 021.
- [40] A. M.-C. So, J. Zhang, and Y. Ye, "On approximating complex quadratic optimization problems via semidefinite programming relaxations," *Math. Program.*, vol. 110, no. 1, pp. 93–110, Jun. 2007.
- [41] M. Grant and S. Boyd. "CVX: MATLAB software for disciplined convex programming version 2.2." Jan. 2020. [Online]. Available: <http://cvxr.com/cvx/>
- [42] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs, recent advances in learning and control (a tribute to M. Vidyasagar)," V. Blondel, S. Boyd, and H. Kimura, Eds. *Lecture Notes in Control and Information Sciences*. Berlin, Germany: Springer-Verlag, 2008, pp. 95–110.
- [43] Löfberg, J., "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. Int. Conf. Robot. Autom. CACSD*, 2004, pp. 284–289.
- [44] *The MOSEK Optimization Toolbox for MATLAB Manual. Version 9.2.* (2021). MOSEK. [Online]. Available: <https://docs.mosek.com/9.2/toolbox/index.html>



TANG CHAO received the B.S. degree in electronics engineering from National Chiao Tung University, Taiwan, in 2019, and the M.S. degree in electronics engineering from National Yang Ming Chiao Tung University, Taiwan, and KU Leuven, Belgium, in 2021. She is currently pursuing the Doctoral degree with the Department of Communications and Information Engineering, Tokyo Institute of Technology, Japan. Her research interests are in the areas of signal processing, machine learning, and data analytics, with a current focus on general learning frameworks for graph signal series utilizing Riemannian manifold geometry.



CARRSON C. FUNG (Member, IEEE) received the B.S. degree in electrical engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1994, the M.S. degree in electrical engineering from Columbia University, New York City, NY, USA, in 1996, and the Ph.D. degree in electrical engineering from The Hong Kong University of Science and Technology in 2005. He was a Member of a Technical Staff with AT&T and Lucent Technologies Bell Laboratories, Holmdel, NJ, USA, from 1994 to 1999, where he worked

on video and audio coding. He was also a Researcher with Hong Kong Applied Science and Technology Research Institute in 2005, where he worked on MIMO-OFDM systems and a Senior DSP Engineer with Sennheiser Research Lab, Palo Alto, CA, USA, in 2006, where he worked on microphone and microphone array technologies. He is currently an Associate Professor with National Yang Ming Chiao Tung University, Hsinchu, Taiwan. His research interests include network and data science, machine learning for signal processing and communications, and optimization. He was the recipient of the prestigious Sir Edward Youde Ph.D. Fellowship from 2001 to 2002. He was the recipient of the National Chiao Tung University College of Electrical and Computer Engineering Excellent Teaching Award in 2011, 2017, and 2018. He received the National Chiao Tung University College of Electrical and Computer Engineering Outstanding Teaching Award in 2020. He also received the National Chiao Tung University Excellent Teaching Award in 2019. He became a Fellow of the Advance Higher Education Academy in 2020.



ZI-EN NI (Student Member, IEEE) received the B.S. degree from the Department of Communication, Navigation and Control Engineering, National Taiwan Ocean University in 2020, and the M.S. degree from the Institute of Electronics, National Yang Ming Chiao Tung University in 2023. His research was focused on wireless communication systems assisted by intelligent reflecting surfaces, mainly joint beamforming design and positioning of IRS.



MYKOLA SERVETNYK (Member, IEEE) received the first B.Sc. degree in electronics engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2015, the second B.Sc. degree in applied mathematics from the Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, in 2017, and the Ph.D. degree in electronics engineering and computer science from National Chiao Tung University in 2020. In 2019, he was a Visiting Research Assistant with the University of Houston, TX, USA. He is currently a software engineer of tests for mobile wireless communications systems and smartphone modem design. His research interests include mobile communications, network planning, distributed algorithm design for machine learning tasks, and data science.