# Angle Aware User Cooperation for Secure Massive MIMO in Rician Fading Channel

Shuai Wang, *Member, IEEE*, Miaowen Wen, *Senior Member, IEEE*,

Minghua Xia, *Member, IEEE*, Rui Wang, *Member, IEEE*,

Qi Hao, *Member, IEEE*, and Yik-Chung Wu, *Senior Member, IEEE*

## Abstract

Massive multiple-input multiple-output communications can achieve high-level security by concentrating radio frequency signals towards the legitimate users. However, this system is vulnerable in a Rician fading environment if the eavesdropper positions itself such that its channel is highly "similar" to the channel of a legitimate user. To address this problem, this paper proposes an angle aware user cooperation (AAUC) scheme, which avoids direct transmission to the attacked user and relies on other users for cooperative relaying. The proposed scheme only requires the eavesdropper's angle information, and adopts an angular secrecy model to represent the average secrecy rate of the attacked system. With this angular model, the AAUC problem turns out to be nonconvex, and a successive convex optimization

algorithm, which converges to a Karush-Kuhn-Tucker solution, is proposed. Furthermore, a closed-form solution and a Bregman first-order method are derived for the cases of large-scale antennas and large-scale users, respectively. Extension to the intelligent reflecting surfaces based scheme is also discussed. Simulation results demonstrate the effectiveness of the proposed successive convex optimization based AAUC scheme, and also validate the low-complexity nature of the proposed large-scale optimization algorithms.

**Index Terms**

Angle aware user cooperation, first-order method, intelligent reflecting surfaces, massive MIMO, physical-layer security, Rician fading.

# I. INTRODUCTION

While 5G is expected to be commercially available in 2020, it cannot fully satisfy the performance requirements of massive connectivity [1]. As a result, beyond 5G aims to serve massive devices with lower latency, higher reliability, and better security. Among the above metrics, security is considered as one of the most challenging problems in wireless communications [2], and it is becoming even more imperative in the era of big data [3], as the massive data usually contains privacy messages such as personal information and control signals. Traditionally, security is enforced by adopting key encryption to prevent the eavesdropper from decoding the message transmitted by legitimate users [4]. However, it is shown by Wyner [5] that secure communication in the presence of an eavesdropper can be guaranteed without any key encryption, as long as the users' channels are better than the eavesdropper's channel [5]–[8]. This is called physical-layer security.

## A. Motivation

In order to achieve the above channel condition for physical-layer security, massive multiple-input multiple-output (MIMO) communication is a promising technique, since massive MIMO can concentrate the radio frequency (RF) signals towards the legitimate users while suppressing the RF power leaked to the eavesdropper [9]–[14]. This leads to a situation that the received signal power at the legitimate user is several orders of magnitude larger than the received signal power at the eavesdropper, thus enabling excellent security without any extra effort [12]–[14]. However, there are still chances for the eavesdropper to overhear the information via active and passive attacks [12], [13]. In an active attack, the eavesdropper adopts pilot contamination to

interfere the channel estimation procedure at the base station (BS) [15], misleading the BS to transmit signals towards the eavesdropper. Such an attack has been recently addressed by careful pilot designs [15]–[17]. Nonetheless, a passive attack is more difficult to deal with, since the passive eavesdropper can hide itself and therefore the BS would have very limited knowledge about the eavesdropper's channel [18]–[21]. If the eavesdropper's channel is independent of users' channels, massive MIMO with artificial noise can be used to prevent eavesdropping [18]–[21]. But if the system operates in a Rician fading environment (i.e., there exists a line-of-sight (LoS) link [22]–[24]), the eavesdropper can position itself such that its channel to the BS is highly "similar" to the channel from a legitimate user to the BS [12, Sec. IV-A], and this passive attack could pose great threats to the massive MIMO system.

## B. Contributions

In this paper, we propose an angle aware user cooperation (AAUC) scheme to combat the above passive attack in massive MIMO systems. The proposed AAUC scheme avoids direct transmission to the attacked user from BS and counts on other users for cooperative relaying. Furthermore, in contrast to existing works (e.g., [21]) that require full channel information of the eavesdropper, our scheme only requires partial channel information of the eavesdropper. In particular, the proposed AAUC scheme exploits the information that the eavesdropper wants to have a similar channel to that of a legitimate user, for which its angle of departure would be similar to that of the attacked user. Based on this angle awareness, the proposed AAUC scheme automatically reduces the transmit powers of users that are close to the potential eavesdropping region, and an optimization problem is formulated to maximize the average secrecy rate subject to the total power constraint.

Nonetheless, the formulated AAUC problem involves numerical integration in the objective function. To circumvent this obstacle, an angular secrecy model, which matches the numerical integration very well, is proposed. Based on the secrecy model, the numerical function is converted into an analytical yet nonconvex function, and a successive convex optimization (SCO) algorithm, which is guaranteed to converge to a Karush-Kuhn-Tucker (KKT) solution to the AAUC problem, is derived. Furthermore, in the large-scale settings, the closed-form solution and the Bregman first-order method (BFOM) are proposed when the number of antennas and the number of users are large, respectively. The two large-scale methods reduce the computation time by orders of magnitude compared to SCO based AAUC. Via integration with intelligent

reflecting surfaces (IRS), the power cost of AAUC can be further reduced. Simulation results demonstrate the effectiveness of the proposed SCO based AAUC, and validate the low-complexity nature of the closed-form AAUC and the BFOM based AAUC algorithms.

*C. Outline*

The rest of this paper is organized as follows. System model and problem formulation are described in Section II and Section III, respectively. The angular secrecy model and the proposed SCO based AAUC scheme are presented in Section IV. The large-scale optimization algorithms are derived in Section V, and extension to the IRS based scheme is discussed in Section VI. Finally, numerical results are presented in Section VII, and conclusions are drawn in Section VIII.

*Notation*: Italic letters, lowercase and uppercase bold letters represent scalars, vectors, and matrices, respectively. The operators $\text{Tr}(\cdot), (\cdot)^T, (\cdot)^H$ and $(\cdot)^{-1}$ take the trace, transpose, Hermitian, and inverse of a matrix, respectively. The operator $[x]^+ = \max(x, 0)$. The symbol $\mathbf{I}_N$ represents the $N \times N$ identity matrix. The symbol $\mathcal{CN}(0, 1)$ represents complex Gaussian distribution with zero mean and unit variance, and $\mathcal{U}(x, y)$ represents the uniform distribution within the interval $[x, y]$. Finally, $\mathbb{E}(\cdot)$ represents the expectation of a random variable, and $\exp(\cdot)$ represents the exponential function of a scalar.

## II. SYSTEM MODEL

We consider a multicast system with LoS links, which consists of a BS with $N$ antennas, $K$ single-antenna legitimate users, and a single-antenna eavesdropper. As shown in Fig. 1, the BS intends to multicast common information to the $K$ legitimate users, while the eavesdropper intends to overhear the signal[1]. In particular, the eavesdropper hides at the line segment between the BS and one of the users, denoted as user $K$. In this way, the eavesdropper can receive a highly correlated signal with that of user $K$ [12], [13], thus decoding the transmitted information. To address this problem, the AAUC scheme consisting of two phases with equal duration is proposed, and the details are given below.

---

[1]The detection of a passive attack is based on spectrum sensing [25].
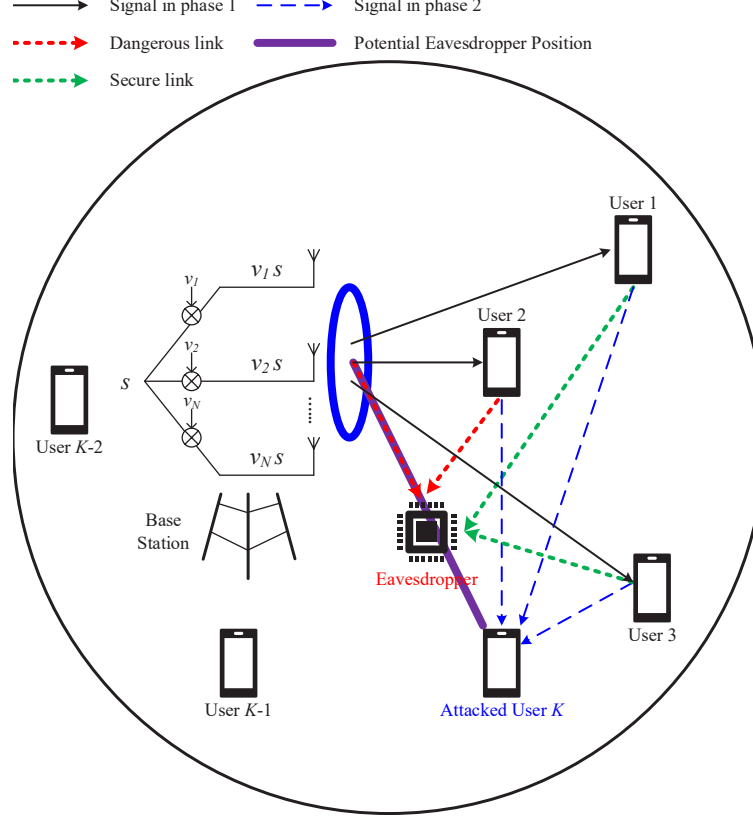
Fig. 1. System model of massive MIMO communication under the LoS attack.

## A. Multicasting Phase

In the multicasting phase, the BS transmits a signal $s$ with $\mathbb{E}[|s|^2] = 1$ to users $\{1, 2, \cdots, K\}$ through a transmit beamforming vector $\mathbf{v} \in \mathbb{C}^{N \times 1}$ with power $||\mathbf{v}||_2^2$. Accordingly, the received signal at a legitimate user $k$ is $\mathbf{g}_k^H \mathbf{v} s + n_k$, where $\mathbf{g}_k \in \mathbb{C}^{N \times 1}$ is the channel vector from the BS to the user $k$, and $n_k \in \mathbb{C}$ is the zero-mean Gaussian noise at the user $k$ with power $\sigma_k^2$. Denoting $D_k$ and $\theta_k$ as the distance and the azimuth angle (i.e., 0 means "north" and $\pi/2$ means "east") between BS and user $k$, the channel $\mathbf{g}_k$ obeys the following Rician fading [26], [27]:

$$
\mathbf{g}_k = \sqrt{\varrho_0 \left( \frac{D_k}{d_0} \right)^{-\alpha}} \left( \sqrt{\frac{K_R}{1 + K_R}} \mathbf{g}_k^{\text{LOS}} \right.
$$
$$
\left. + \sqrt{\frac{1}{1 + K_R}} \mathbf{g}_k^{\text{NLOS}} \right), \quad k = 1, \cdots, K, \tag{1}
$$

where $\varrho_0$ is the pathloss at distance $d_0 = 1\,\text{m}$, and $\alpha$ is the pathloss exponent. Notice that $K_R$ is the Rician K-factor accounting for propagation effects of the LoS and non-LoS links, which

can be pre-determined from a few channel measurements in the environment [28]. Furthermore, the LoS component $\mathbf{g}_k^{\mathrm{LOS}}$ is [29]

$$\mathbf{g}_k^{\mathrm{LOS}} = \Big[ 1, \exp\left(-\mathrm{j}\pi \sin\theta_k\right), \cdots ,$$
$$\exp\left(-(N-1)\,\mathrm{j}\pi \sin\theta_k\right) \Big]^T, \tag{2}$$

and the non-LoS component $\mathbf{g}_k^{\mathrm{NLOS}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$.

Since the angle of eavesdropper $\theta_E$ can be approximated as $\theta_E \approx \theta_K$, $\mathbf{g}_E^{\mathrm{LOS}} \approx \mathbf{g}_K^{\mathrm{LOS}}$ holds, where $\mathbf{g}_E^{\mathrm{LOS}}$ is the LoS component of the channel $\mathbf{g}_E \in \mathbb{C}^{N \times 1}$ from the BS to the eavesdropper. According to $\mathbf{g}_E^{\mathrm{LOS}} \approx \mathbf{g}_K^{\mathrm{LOS}}$ and (1), we have $\mathbf{g}_E // \mathbf{g}_K$ due to $K_R \gg 1$ in LoS environment, where $\mathbf{a} // \mathbf{b}$ means that vectors $\mathbf{a}$ and $\mathbf{b}$ have the same direction. To ensure that $|\mathbf{g}_E^H \mathbf{v}| = 0$, the BS needs to design $\mathbf{v}$ such that $|\mathbf{g}_K^H \mathbf{v}| = 0$. Therefore, the achievable rates at the user $K$ and the eavesdropper can be approximated to zero. On the other hand, the achievable rate $R_k$ at user $k \neq K$ is

$$R_k = \frac{1}{2}\log_2\left(1 + \frac{|\mathbf{g}_k^H \mathbf{v}|^2}{\sigma_k^2}\right), \quad \forall k \neq K, \tag{3}$$

where the factor $1/2$ is due to the two transmission phases.

### B. User Cooperation Phase

Due to $|\mathbf{g}_K^H \mathbf{v}| = 0$, the user $K$ would hardly receive information during the multicasting phase. As a result, the other users need to help to forward the information to user $K$. In particular, the helping users $\{1, \cdots, K-1\}$ transmit the signal $s$ to the user $K$ through a beamforming vector $\mathbf{w} \in \mathbb{C}^{(K-1) \times 1} = [w_1, ..., w_{K-1}]^T$ in the user cooperation phase, and the received signal at the user $K$ is given by $\mathbf{h}_K^H \mathbf{w} s + n_K$, where $\mathbf{h}_K \in \mathbb{C}^{(K-1) \times 1}$ is the channel vector[2] from the helping users to the user $K$. Therefore, the achievable rate $R_K$ at user $K$ is [30]

$$R_K = \min\left\{ R_1, \cdots, R_{K-1}, \frac{1}{2}\log_2\left(1 + \frac{|\mathbf{h}_K^H \mathbf{w}|^2}{\sigma_K^2}\right) \right\}. \tag{4}$$

On the other hand, the received signal at the eavesdropper is given by $\mathbf{h}_E^H \mathbf{w} s + n_E$, where $\mathbf{h}_E \in \mathbb{C}^{(K-1) \times 1}$ is the channel vector from the helping users to the eavesdropper and $n_E \in \mathbb{C}$ is

---

[2]Channel $\mathbf{h}_K$ can be estimated at the user $K$, who subsequently forwards $\mathbf{h}_K$ to the BS.

the zero-mean Gaussian noise at the eavesdropper with power $\sigma_E^2$. Accordingly, the eavesdropping rate is[3]

$$R_E = \min\left\{R_1, \cdots, R_{K-1}, \frac{1}{2}\log_2\left(1 + \frac{|\mathbf{h}_E^H\mathbf{w}|^2}{\sigma_E^2}\right)\right\}. \tag{5}$$

Finally, combining the results in (3)–(5), the secrecy rate of the multicast system under the AAUC scheme can be expressed as [5]–[7]:

$$\left[\min_{k=1,\cdots,K} R_k - R_E\right]^+ = \left[R - \frac{1}{2}\log_2\left(1 + \frac{|\mathbf{h}_E^H\mathbf{w}|^2}{\sigma_E^2}\right)\right]^+, \tag{6}$$

where the equality is due to $R_K \leq R_k$ for any $k \neq K$ according to (4), and

$$R = \frac{1}{2}\min\left\{\log_2\left(1 + \frac{|\mathbf{g}_1^H\mathbf{v}|^2}{\sigma_1^2}\right), \cdots, \right.$$

$$\left. \log_2\left(1 + \frac{|\mathbf{g}_{K-1}^H\mathbf{v}|^2}{\sigma_{K-1}^2}\right), \log_2\left(1 + \frac{|\mathbf{h}_K^H\mathbf{w}|^2}{\sigma_K^2}\right)\right\}. \tag{7}$$

## III. PROBLEM FORMULATION

### A. Average Secrecy Rate

The major challenge to maximize (6) is that $\mathbf{h}_E$ is unknown. A traditional way is to model $\mathbf{h}_E$ as a constant value $\widehat{\mathbf{h}}_E$ plus some random perturbations [21], [31]. However, such a method requires the knowledge of $\widehat{\mathbf{h}}_E$, which is difficult to obtain since the eavesdropper is passive. To combat the passive attack without knowing the full channel information, we propose to exploit the angular information. In particular, denoting the location of BS as $(0, 0)$, the location of user $k$ can be expressed as $(D_k\cos\theta_k, D_k\sin\theta_k)$, where $D_k\cos\theta_k$ is the latitude and $D_k\sin\theta_k$ is the longitude. Since the eavesdropper hides at the line segment between BS and user $K$, the location of eavesdropper can be represented as $(\rho\cos\theta_K, \rho\sin\theta_K)$, where $\rho$ is bounded from $0$ to $D_K$.

Based on the locations, the distance from user $k$ to eavesdropper is given by

$$d_{E,k}(\rho) = \left[(\rho\cos\theta_K - D_k\cos\theta_k)^2 \right.$$

$$\left. + (\rho\sin\theta_K - D_k\sin\theta_k)^2\right]^{-1/2}. \tag{8}$$

---

[3]In practice, the eavesdropper may not know $\mathbf{w}$. However, the eavesdropper can adjust its receiver to fully exploit the received signal strength $|\mathbf{h}_E^H\mathbf{w}|^2$.

By further adopting the Rician fading channel model [26], [27], the channel $h_{E,k}$ (which is the $k^{\text{th}}$ element in $\mathbf{h}_E = [h_{E,1}, \cdots, h_{E,K-1}]^T$) from user $k$ to eavesdropper can be modeled as:

$$h_{E,k}(\rho, \phi_k, \delta_k) = \sqrt{\varrho_0 \left(\frac{d_{E,k}(\rho)}{d_0}\right)^{-\alpha}}$$

$$\times \left(\sqrt{\frac{K_R}{1+K_R}} \, \mathrm{e}^{\mathrm{j}\phi_k} + \sqrt{\frac{1}{1+K_R}} \, \delta_k\right), \tag{9}$$

where $\phi_k$ is the phase[4] of LoS link with $\phi_k \sim \mathcal{U}(-\pi, +\pi)$, and $\delta_k$ is the non-LoS component with $\delta_k \sim \mathcal{CN}(0, 1)$.

By substituting the channel model (9) into (6), the secrecy rate for a fixed $(\rho, \{\phi_k, \delta_k\})$ can be obtained. Then by averaging the result in (6) over $(\rho, \{\phi_k, \delta_k\})$, the average secrecy rate of the system can be computed as follows [32]–[34]:

$$S(R, \mathbf{w}|\sigma_E) = \frac{1}{D_K} \int_0^{D_K} \mathbb{E}_{\{\phi_k, \delta_k\}} \left\{ \left[ R \right. \right.$$

$$\left. \left. - \frac{1}{2}\log_2\left(1 + \frac{\left|\mathbf{h}_E^H(\rho, \{\phi_k, \delta_k\})\mathbf{w}\right|^2}{\sigma_E^2}\right) \right]^+ \right\} \mathrm{d}\rho. \tag{10}$$

### B. Secrecy Rate Maximization

In the considered system, the design variables that can be controlled are the transmit beamformer $\mathbf{v}$ at BS and the distributed beamformer $\mathbf{w}$ at helping users. Since the power costs at both BS and helping users should be smaller than the total budget, the beamformers need to satisfy $\frac{\|\mathbf{v}\|^2}{2} + \frac{\|\mathbf{w}\|^2}{2} \leq P_{\max}$, where the factor $1/2$ is due to the two transmission phases, and $P_{\max}$ is the total transmit power budget. Having the transmit power satisfied, it is then crucial to maximize the average secrecy rate $S(R, \mathbf{w}|\sigma_E)$, which leads to the following optimization problem:

$$\text{P1}: \max_{\mathbf{v}, \mathbf{w}, R} \ \frac{1}{D_K} \int_0^{D_K} \mathbb{E}_{\{\phi_k, \delta_k\}} \left\{ \left[ R - \frac{1}{2}\log_2\left(1 \right. \right. \right.$$

$$\left. \left. \left. + \frac{\left|\mathbf{h}_E^H(\rho, \{\phi_k, \delta_k\})\mathbf{w}\right|^2}{\sigma_E^2}\right) \right]^+ \right\} \mathrm{d}\rho,$$

$$\text{s.t.} \quad \frac{\|\mathbf{v}\|_2^2}{2} + \frac{\|\mathbf{w}\|_2^2}{2} \leq P_{\max}, \tag{11a}$$

$$|\mathbf{g}_K^H \mathbf{v}| = 0, \tag{11b}$$

---

[4]Since each user has its own local phase shift in the transmitted signal, $\phi_k$ is a random value independent of $\rho$.

$$R \leq \frac{1}{2}\log_2\left(1 + \frac{|\mathbf{g}_k^H\mathbf{v}|^2}{\sigma_k^2}\right), \quad \forall k \neq K, \tag{11c}$$

$$R \leq \frac{1}{2}\log_2\left(1 + \frac{|\mathbf{h}_K^H\mathbf{w}|^2}{\sigma_K^2}\right), \tag{11d}$$

where the objective function of P1 is obtained from (10). The constraint (11a) is the operation budget constraint and the constraint (11b) guarantees that no information is leaked to the eavesdropper during the multicasting phase. Finally, the constraints (11c) and (11d) are the data-rate constraints obtained from (7).

Notice that problem P1 can be extended to the case with different information signals for different users by adding an additional individual transmission phase for the users $\{1, \cdots, K-1\}$. In addition, the problem P1 can be extended to the case with separate transmit power constraints at BS and users. In such a case, the design variables $\mathbf{v}$ and $\mathbf{w}$ are no longer coupled in the constraints, and the proposed algorithms are still applicable to the resultant problem.

## IV. PROPOSED SCO BASED AAUC

### A. Angular Secrecy Model

To solve P1, the first challenge comes from the integration over $\rho$ and the expectation over $\{\phi_k, \delta_k\}$, which make the objective a numerical function of $\mathbf{w}$. To address this challenge, it is observed from (10) that the secrecy rate $S$ satisfies the following properties:

(i) When $\sigma_E^2 \to 0$, we must have $S \to 0$.

(ii) When $\sigma_E^2 \to +\infty$, the secrecy rate $S$ would reach its maximum value $R$.

(iii) The secrecy rate $S$ is lower bounded by a logarithm function $\left[R - \frac{1}{2}\log_2\left(1 + \mathbf{w}^H\mathbf{J}\mathbf{w}/\sigma_E^2\right)\right]^+$
    (proved in Appendix A), where $\mathbf{J} = \mathrm{diag}\left(J_1, \cdots, J_{K-1}\right)$ is the angular matrix with

$$J_k = \frac{\varrho_0}{D_k}\int_0^{D_K}\Big[(\rho\cos\theta_K - D_k\cos\theta_k)^2$$
$$+ (\rho\sin\theta_K - D_k\sin\theta_k)^2\Big]^{-\alpha/2}\mathrm{d}\rho. \tag{12}$$

Based on (i)–(iii), the average secrecy rate should have a logarithm curve with respect to noise power $\sigma_E^2$. To this end, we propose an angular secrecy model as follows:

$$\widehat{S}\left(R, \mathbf{w}|\sigma_E, \lambda\right) = \left[R - \frac{1}{2}\log_2\left(1 + \frac{\lambda\mathbf{w}^H\mathbf{J}\mathbf{w}}{\sigma_E^2}\right)\right]^+, \tag{13}$$

where $\lambda \in [0, 1]$ is a tuning parameter. It can be verified that the nonlinear function in (13) satisfies all the properties (i)–(iii). Moreover, since equation (12) does not involve $\mathbf{w}$ inside the integral, the angular matrix $\mathbf{J}$ can be computed via one-dimensional integration.
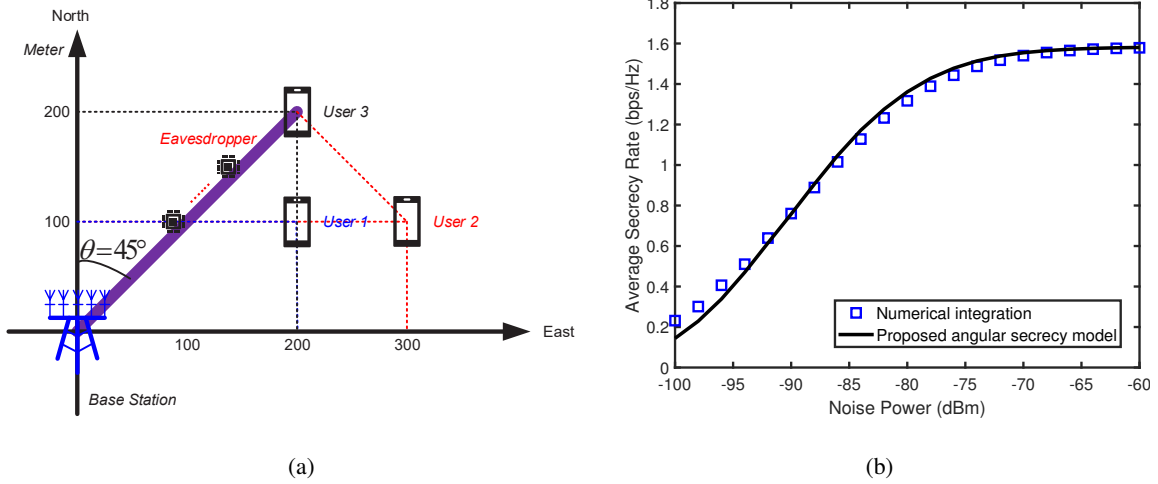
Fig. 2. a) Illustrating example with $K = 3$, $\alpha = 2.5$ and $K_R = 30$ dB; b) Fitting performance of the angular secrecy model.

**Interpretation of the Angular Matrix.** $J_k$ in (12) is the average pathloss from user $k$ to the eavesdropper and quantitatively measures how dangerous for user $k$ to transmit a signal. For example, in Fig. 2a with $K = 3$ and $\alpha = 2.5$, we have $J_1 = 10^{-5}\varrho_0$ and $J_2 = 2 \times 10^{-6}\varrho_0$ according to (12), meaning that user 2 can provide much better security. This corroborates the fact that user 1 is closer to the line segment between BS and user 3.

Based on (i)–(iii), $\widehat{S}_\lambda(R, \mathbf{w}|\sigma_E, \lambda)$ would have a similar trend to $S$, and the next step is to minimize the gap between the two curves $\widehat{S}(R, \mathbf{w}|\sigma_E, \lambda)$ and $S(R, \mathbf{w}|\sigma_E)$. Since $\widehat{S}$ is a decreasing function of $\lambda$, one-dimensional search over $[0, 1]$ can be adopted to determine an appropriate $\lambda$. More specifically, we numerically compute $S(R^{(i)}, \mathbf{w}^{(i)}|\sigma_E^{(i)})$ using (10) for the input $(R^{(i)}, \mathbf{w}^{(i)}, \sigma_E^{(i)})$, where $i = 1, \cdots, T$ is the index of training sample and $T$ is the sample size. With the training set $\{(R^{(i)}, \mathbf{w}^{(i)}, \sigma_E^{(i)})\}_{i=1}^{T}$, the parameter $\lambda$ in $\widehat{S}$ can be found via nonlinear least squares fitting:

$$\min_{\lambda \in [0,1]} \ \frac{1}{T} \sum_{i=1}^{T} \left| S(R^{(i)}, \mathbf{w}^{(i)}|\sigma_E^{(i)}) - \widehat{S}\left(R^{(i)}, \mathbf{w}^{(i)}|\sigma_E^{(i)}, \lambda\right) \right|^2. \tag{14}$$

In order to verify the accuracy of the model, we consider the setting of Fig. 2a. In this environment, we consider 41 different values of $[\sigma_E^{(i)}]^2$ ranging from $-100\,\mathrm{dBm}$ to $-60\,\mathrm{dBm}$ by a step size of $2\,\mathrm{dBm}$, and $T = 21 \times 100$ training samples are generated with $R^{(i)} \sim \mathcal{U}(0, 3)$ and $\mathbf{w}^{(i)} \sim \mathcal{CN}(\mathbf{0}, \frac{0.01}{K-1}\mathbf{I}_{K-1})$ (i.e., the average power of $\mathbf{w}^{(i)}$ is $10\,\mathrm{dBm}$). It can be seen from Fig. 2b that with the choice of $\lambda = 0.64$, the proposed model matches the numerical simulation of

$S$ very well under a large range of noise power $\sigma_E^2$. Moreover, the mean square error of (14) at $\lambda = 0.64$ is $0.0056$, which is negligible compared to the absolute secrecy rate.

*Remark 1:* Notice that the proposed method can be well extended to the 3D scenarios when the eavesdropper is not on the same plane as the users (e.g., the eavesdropper is a flying unmanned aerial vehicle [35]). In such a case, the distance from eavesdropper to the user $k$ is

$$d_{E,k}(\rho) = \Big[(\rho \cos \theta_K - D_k \cos \theta_k)^2$$
$$+ (\rho \sin \theta_K - D_k \sin \theta_k)^2 + \rho^2 \tan^2 \beta \Big]^{-1/2}, \tag{15}$$

where $\beta$ is the elevation angle of the eavesdropper. Accordingly, the $k^{\text{th}}$ element of the angular matrix becomes

$$J_k = \frac{\varrho_0}{D_K} \int_0^{D_K} \Big[(\rho \cos \theta_K - D_k \cos \theta_k)^2$$
$$+ (\rho \sin \theta_K - D_k \sin \theta_k)^2 + \rho^2 \tan^2 \beta \Big]^{-\alpha/2} d\rho. \tag{16}$$

Once the angular matrix is defined, all the derivations in the manuscript stay the same for 2D and 3D scenarios.

### B. Proposed SCO Algorithm

With the angular secrecy model, we can replace $S(R, \mathbf{w}|\sigma_E)$ with $\widehat{S}(R, \mathbf{w}|\sigma_E, \lambda)$ in the objective function of P1, and the resultant problem is given by

$$\max_{\mathbf{v}, \mathbf{w}, R} \widehat{S}(R, \mathbf{w}|\sigma_E, \lambda) \quad \text{s.t.} \quad (11a) - (11d). \tag{17}$$

To solve (17), the following transformations are used to eliminate the constraints.

- QR decomposition is applied to $\mathbf{g}_K$, and the unitary matrix $\mathbf{U} \in \mathbb{C}^{N \times (N-1)}$, which spans the null space of $\mathbf{g}_K$, is obtained. With $\mathbf{U}$, we can set $\mathbf{v} = \mathbf{U}\mathbf{z}$, where $\mathbf{z} \in \mathbb{C}^{(N-1) \times 1}$ is a newly introduced variable.

- Putting $\mathbf{v} = \mathbf{U}\mathbf{z}$ into (11a), this constraint becomes $||\mathbf{z}||_2^2 + ||\mathbf{w}||_2^2 \leq 2P_{\max}$, where we have used $\mathbf{U}^H \mathbf{U} = \mathbf{I}$.

- Putting $\mathbf{v} = \mathbf{U}\mathbf{z}$ into (11b), we have $|\mathbf{g}_K^H \mathbf{U}\mathbf{z}| = 0$. This constraint is always feasible due to $\mathbf{g}_K^H \mathbf{U} = \mathbf{0}^T$.

- Putting (11c) and (11d) into the objective function of (17), the inequality constraints (11c)–(11d) can be removed.

Based on the above procedure, problem (17) is equivalently reformulated into

$$\text{P2}: \max_{\mathbf{z},\mathbf{w}} \min \left[\min_{k \neq K} \ \Phi_k(\mathbf{z},\mathbf{w}), \Upsilon(\mathbf{w})\right],$$

$$\text{s.t.} \ \ ||\mathbf{z}||_2^2 + ||\mathbf{w}||_2^2 \leq 2P_{\max}, \tag{18}$$

where

$$\Phi_k(\mathbf{z},\mathbf{w}) = \frac{1}{2}\log_2\left(1 + \frac{|\mathbf{g}_k^H \mathbf{U}\mathbf{z}|^2}{\sigma_k^2}\right)$$

$$- \frac{1}{2}\log_2\left(1 + \frac{\lambda \mathbf{w}^H \mathbf{J}\mathbf{w}}{\sigma_E^2}\right), \tag{19}$$

$$\Upsilon(\mathbf{w}) = \frac{1}{2}\log_2\left(1 + \frac{|\mathbf{h}_K^H \mathbf{w}|^2}{\sigma_K^2}\right)$$

$$- \frac{1}{2}\log_2\left(1 + \frac{\lambda \mathbf{w}^H \mathbf{J}\mathbf{w}}{\sigma_E^2}\right). \tag{20}$$

Due to the non-concave functions $\Phi_k$ and $\Upsilon$, problem P2 is nonconvex. To address this nonconvexity, we will propose a successive convex optimization (SCO) algorithm [36]–[41], which constructs a sequence of lower bounds on $\{\Phi_k, \Upsilon\}$ and solves the surrogate problems. There are many techniques to construct the surrogate functions, e.g., convexity inequality, second-order Taylor expansion, arithmetic-geometric mean inequality, and Cauchy-Schwartz inequality [36]. In this paper, the surrogate functions are found by first-order Taylor expansion.

More specifically, given any feasible solution $(\mathbf{z}^\star, \mathbf{w}^\star)$ to P2, we define surrogate functions

$$\widetilde{\Phi}_k(\mathbf{z},\mathbf{w}|\mathbf{z}^\star,\mathbf{w}^\star)$$

$$= \frac{1}{2}\log_2\left(1 + 2\text{Re}\left[\frac{(\mathbf{z}^\star)^H \mathbf{U}^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{U}\mathbf{z}}{\sigma_k^2}\right] - \frac{|\mathbf{g}_k^H \mathbf{U}\mathbf{z}^\star|^2}{\sigma_k^2}\right)$$

$$- \frac{\lambda \mathbf{w}^H \mathbf{J}\mathbf{w} - \lambda (\mathbf{w}^\star)^H \mathbf{J}\mathbf{w}^\star}{2\ln 2 \left[\sigma_E^2 + \lambda (\mathbf{w}^\star)^H \mathbf{J}\mathbf{w}^\star\right]}$$

$$- \frac{1}{2}\log_2\left(1 + \frac{\lambda (\mathbf{w}^\star)^H \mathbf{J}\mathbf{w}^\star}{\sigma_E^2}\right), \tag{21}$$

$$\widetilde{\Upsilon}(\mathbf{w}|\mathbf{w}^\star)$$

$$= \frac{1}{2}\log_2\left(1 + 2\text{Re}\left[\frac{(\mathbf{w}^\star)^H \mathbf{h}_K \mathbf{h}_K^H \mathbf{w}}{\sigma_K^2}\right] - \frac{|\mathbf{h}_K^H \mathbf{w}^\star|^2}{\sigma_K^2}\right)$$

$$- \frac{\lambda \mathbf{w}^H \mathbf{J}\mathbf{w} - \lambda (\mathbf{w}^\star)^H \mathbf{J}\mathbf{w}^\star}{2\ln 2 \left[\sigma_E^2 + \lambda (\mathbf{w}^\star)^H \mathbf{J}\mathbf{w}^\star\right]}$$

$$-\frac{1}{2}\log_2\left(1+\frac{\lambda\left(\mathbf{w}^\star\right)^H\mathbf{J}\mathbf{w}^\star}{\sigma_E^2}\right),\tag{22}$$

and the following proposition can be established.

**Proposition 1.** *The functions* $\{\widetilde{\Phi}_k,\widetilde{\Upsilon}\}$ *satisfy the following properties: (i)* $\widetilde{\Phi}_k(\mathbf{z},\mathbf{w}|\mathbf{z}^\star,\mathbf{w}^\star)\le$ $\Phi_k(\mathbf{z},\mathbf{w})$ *and* $\widetilde{\Upsilon}(\mathbf{w}|\mathbf{w}^\star)\le\Upsilon(\mathbf{w})$*; (ii)* $\widetilde{\Phi}_k(\mathbf{z}^\star,\mathbf{w}^\star|\mathbf{z}^\star,\mathbf{w}^\star)=\Phi_k(\mathbf{z}^\star,\mathbf{w}^\star)$ *and* $\widetilde{\Upsilon}(\mathbf{w}^\star|\mathbf{w}^\star)=\Upsilon(\mathbf{w}^\star)$*;* *(iii) With complex gradient operator* $\nabla_{\mathbf{x}}:=\partial/\partial\operatorname{conj}(\mathbf{x})$*,*

$$\nabla_{\mathbf{z}}\widetilde{\Phi}_k(\mathbf{z},\mathbf{w}|\mathbf{z}^\star,\mathbf{w}^\star)\Big|_{\mathbf{z}=\mathbf{z}^\star,\mathbf{w}=\mathbf{w}^\star}=\nabla_{\mathbf{z}}\Phi_k(\mathbf{z},\mathbf{w})\Big|_{\mathbf{z}=\mathbf{z}^\star,\mathbf{w}=\mathbf{w}^\star},$$

$$\nabla_{\mathbf{w}}\widetilde{\Phi}_k(\mathbf{z},\mathbf{w}|\mathbf{z}^\star,\mathbf{w}^\star)\Big|_{\mathbf{z}=\mathbf{z}^\star,\mathbf{w}=\mathbf{w}^\star}=\nabla_{\mathbf{w}}\Phi_k(\mathbf{z},\mathbf{w})\Big|_{\mathbf{z}=\mathbf{z}^\star,\mathbf{w}=\mathbf{w}^\star},$$

$$\nabla_{\mathbf{w}}\widetilde{\Upsilon}(\mathbf{w}|\mathbf{w}^\star)\Big|_{\mathbf{w}=\mathbf{w}^\star}=\nabla_{\mathbf{w}}\Upsilon(\mathbf{w})\Big|_{\mathbf{w}=\mathbf{w}^\star}.$$

*Proof.* See Appendix B. ■

With the above observation, a lower bound can be directly obtained if we replace the functions $\{\Phi_k,\Upsilon\}$ by $\{\widetilde{\Phi}_k,\widetilde{\Upsilon}\}$ expanded around a feasible point. However, a tighter lower bound can be achieved if we treat the obtained solution as another feasible point and continue to construct the next-round surrogate function. In particular, assuming that the solution at the $n^{\text{th}}$ iteration is given by $(\mathbf{z}^{[n]},\mathbf{w}^{[n]})$, the proposed SCO executes the following two steps at the $(n+1)^{\text{th}}$ iteration:

- Use CVX Mosek [42], a Matlab software package for convex optimization, to solve

$$\text{P2}[n+1]:\max_{\mathbf{z},\mathbf{w}}\min\left[\min_{k\ne K}\ \widetilde{\Phi}_k(\mathbf{z},\mathbf{w}|\mathbf{z}^{[n]},\mathbf{w}^{[n]}),\right.$$
$$\left.\widetilde{\Upsilon}(\mathbf{w}|\mathbf{w}^{[n]})\right],$$
$$\text{s.t.}\ \ ||\mathbf{z}||_2^2+||\mathbf{w}||_2^2\le 2P_{\max}.\tag{23}$$

- Denoting the optimal solution to P2$[n+1]$ as $(\mathbf{z}^*,\mathbf{w}^*)$, we set $(\mathbf{z}^{[n+1]},\mathbf{w}^{[n+1]})=(\mathbf{z}^*,\mathbf{w}^*)$, and the process repeats with solving the problem P2$[n+2]$.

According to **Proposition 1** and [36, Theorem 1], the sequence $\{(\mathbf{z}^{[0]},\mathbf{w}^{[0]}),(\mathbf{z}^{[1]},\mathbf{w}^{[1]}),\cdots\}$ converges to the KKT solution to P2 for any feasible starting point $(\mathbf{z}^{[0]},\mathbf{w}^{[0]})$. In terms of computational complexity, P2$[n+1]$ involves $N+K-2$ variables, and the complexity for solving P2$[n+1]$ can be computed to be $\mathcal{O}[(N+K-2)^{3.5}]$ [43]. As a consequence, the total complexity for solving P2 is $\mathcal{O}[\mathcal{I}(N+K-2)^{3.5}]$, where $\mathcal{I}$ is the number of iterations needed for the SCO based AUCC algorithm to converge.

## V. Large-Scale Optimization Algorithms

While a KKT solution to P2 has been obtained in Section IV, the corresponding SCO algorithm requires a cubic complexity with respect to $N$ and $K$. This leads to extremely time-consuming computations in the case of large $N$ or $K$. Therefore, efficient large-scale optimization algorithms become indispensable. Below, we will sequentially consider the cases of large $N$, large $K$ and large $(N, K)$.

### A. Large Number of Antennas

When $N$ is very large such that $N \gg K$, the channels from BS to users would be approximately orthogonal and we have $|\mathbf{g}_j^H \mathbf{g}_k|^2 / ||\mathbf{g}_k||_2^2 \to 0$ for any $j \neq k$ [9]. Based on such orthogonal feature, the following proposition can be established.

**Proposition 2.** *The asymptotic optimal* $\mathbf{z}^*$ *and* $\mathbf{w}^*$ *to* P2 *when* $N \to +\infty$ *satisfy*

$$\mathbf{z}^* = \sum_{k=1}^{K-1} \sqrt{\frac{(2P_{\max} - ||\mathbf{w}^*||_2^2)\,\sigma_k^2}{\sum_{i=1}^{K-1} \sigma_i^2 / ||\mathbf{U}^H \mathbf{g}_i||_2^2}} \frac{\mathbf{U}^H \mathbf{g}_k}{||\mathbf{U}^H \mathbf{g}_k||_2^2}. \tag{24}$$

*Proof.* See Appendix C. ∎

Using the result from **Proposition 2**, we can eliminate the vector $\mathbf{z}$ in P2 by setting $\mathbf{z} = \mathbf{z}^*$, which would not change the optimal objective value of P2. After the above procedure, $\Phi_k$ in the objective function becomes

$$\Phi_1(\mathbf{z}^*, \mathbf{w}) = \cdots = \Phi_{K-1}(\mathbf{z}^*, \mathbf{w}) = \Psi(\mathbf{w}),$$

where

$$\begin{aligned}
\Psi(\mathbf{w}) = {} & \frac{1}{2}\log_2\left(1 + \frac{2P_{\max} - ||\mathbf{w}||_2^2}{\sum_{i=1}^{K-1} \sigma_i^2 / ||\mathbf{U}^H \mathbf{g}_i||_2^2}\right) \\
& - \frac{1}{2}\log_2\left(1 + \frac{\lambda \mathbf{w}^H \mathbf{J} \mathbf{w}}{\sigma_E^2}\right).
\end{aligned} \tag{25}$$

Moreover, since the constraint (18) is always satisfied by $\mathbf{z}$ in (24), it can be removed accordingly.

Based on the above transformation, problem P2 under $N \to +\infty$ is equivalent to

$$\text{P3}: \max_{\mathbf{w}} \min\left[\Psi(\mathbf{w}), \Upsilon(\mathbf{w})\right]. \tag{26}$$

Now, it can be seen that problem P3 only involves one variable $\mathbf{w}$, and is of much lower dimension than P2. Although P3 is a nonconvex problem, the following proposition is established, which derives the closed-form solution of $\mathbf{w}$.

**Proposition 3.** *The optimal $\mathbf{w}^*$ to P3 is*

$$\mathbf{w}^* = \frac{\left(\boldsymbol{\Xi} + \lambda \mathbf{J}/\sigma_E^2\right)^{-1/2} \mathbf{q}}{\sqrt{\mathbf{q}^H \left(\boldsymbol{\Xi} + \lambda \mathbf{J}/\sigma_E^2\right)^{-1/2} \boldsymbol{\Xi} \left(\boldsymbol{\Xi} + \lambda \mathbf{J}/\sigma_E^2\right)^{-1/2} \mathbf{q}}}, \tag{27}$$

*where*

$$\boldsymbol{\Xi} = \frac{1}{2P_{\max}} \left[ \frac{\mathbf{h}_K \mathbf{h}_K^H}{\sigma_K^2} \left( \sum_{i=1}^{K-1} \frac{\sigma_i^2}{||\mathbf{U}^H \mathbf{g}_i||_2^2} \right) + \mathbf{I} \right], \tag{28}$$

*and $\mathbf{q}$ is the dominant eigenvector of*

$$\left(\boldsymbol{\Xi} + \lambda \mathbf{J}/\sigma_E^2\right)^{-1/2} \left(\boldsymbol{\Xi} + \mathbf{h}_K \mathbf{h}_K^H/\sigma_K^2\right) \left(\boldsymbol{\Xi} + \lambda \mathbf{J}/\sigma_E^2\right)^{-1/2}.$$

*Proof.* See Appendix D. ∎

Since P3 is equivalent to P2 under $N \to \infty$, the asymptotic optimal $\mathbf{w}^*$ to P2 is given by (27) according to **Proposition 3**. By putting $\mathbf{w} = \mathbf{w}^*$ into (24) of **Proposition 2**, the optimal $\mathbf{z}^*$ to P2 under $N \to \infty$ is also obtained. Therefore, the entire procedure for deriving the solution to P2 with large $N$ is to execute (27) and (24) sequentially. The computational complexity of the closed-form AAUC algorithm is dominated by computing $\left(\boldsymbol{\Xi} + \lambda \mathbf{J}/\sigma_E^2\right)^{-1/2}$, which requires a complexity of $\mathcal{O}((K-1)^3)$.

*B. Large Number of Users*

When $K$ is very large, the orthogonal condition does not hold and the closed-form AAUC scheme is not applicable. To this end, this subsection will propose the BFOM based AAUC scheme for solving P2 when $K$ is large.

More specifically, as $K \to +\infty$, $||\mathbf{h}_K||_2^2 \gg \mathrm{Tr}(\mathbf{J})$ holds. Therefore, we have $\Upsilon(\mathbf{w}) \to \frac{1}{2}\log_2\left(1 + |\mathbf{h}_K^H \mathbf{w}|^2/\sigma_K^2\right)$, and the optimal $\mathbf{w}^*$ is given by $\mathbf{w}^* = \sqrt{p}\,\mathbf{h}_K/||\mathbf{h}_K||_2$, with $p \geq 0$ being the transmit power to be determined. By putting $\mathbf{w} = \sqrt{p}\,\mathbf{h}_K/||\mathbf{h}_K||_2$ into P2, P2 under $K \to +\infty$ is re-written as

$$\text{P4}: \max_{\mathbf{z},\, p \geq 0} \ \min \left( \min_{k=1,\ldots,K-1} \frac{|\mathbf{g}_k^H \mathbf{U} \mathbf{z}|^2}{\sigma_k^2}, \frac{p||\mathbf{h}_K||_2^2}{\sigma_K^2} \right),$$

$$\text{s.t.} \ \ ||\mathbf{z}||_2^2 + p \leq 2P_{\max}, \tag{29}$$

where we have removed $\frac{1}{2}\log_2(1+\cdot)$ in the objective function due to its monotonicity. It can be seen from P4 that the objective function to be maximized is non-concave. Therefore, P4 is a

nonconvex optimization problem, which can also be solved via the SCO algorithm. Specifically, with any feasible starting point $(\mathbf{z}^{[0]}, p^{[0]})$, we execute the following update at the $n^{\text{th}}$ iteration:

$$
\begin{aligned}
\left(\mathbf{z}^{[n+1]}, p^{[n+1]}\right) = \underset{\substack{\mathbf{z},\, p \geq 0 \\ ||\mathbf{z}||_2^2 + p \leq 2P_{\max}}}{\operatorname{argmax}} \quad \min \left\{ \min_{k=1,\ldots,K-1} \right. \\
\left. \left(2\operatorname{Re}\left[\frac{(\mathbf{z}^{[n]})^H \mathbf{U}^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{U}\mathbf{z}}{\sigma_k^2}\right] - \frac{|\mathbf{g}_k^H \mathbf{U}\mathbf{z}^{[n]}|^2}{\sigma_k^2}\right), \frac{p||\mathbf{h}_K||_2^2}{\sigma_K^2} \right\},
\end{aligned} \tag{30}
$$

According to [36]–[41], the above iterative procedure is guaranteed to converge to a KKT solution to P4.

Based on the SCO framework, the remaining question is how to compute (30). The major challenge comes from the non-smooth operator $\min_{k=1,\ldots,K-1}$ in the objective function, which hinders us from computing the gradients. To this end, in the following, problem (30) will be reformulated into a smooth bilevel optimization problem with $\ell_1$ norm and $\ell_2$ norm constraints. Furthermore, since the projection onto $\ell_1$ norm in Euclidean space involves a high computational complexity, we propose to execute the projection in non-Euclidean space. This is achieved by using Kullback-Leibler (KL) divergence as the distance measure, and the resultant algorithm is termed BFOM based AAUC.

First of all, we derive the procedure that equivalently transforms (30) into a smooth optimization problem. To this end, variables $\mathbf{x} = [\operatorname{Re}(\mathbf{z}^T), \operatorname{Im}(\mathbf{z}^T)]^T \in \mathbb{R}^{(2N-2)\times 1}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{K \times 1}$ are introduced with $\mathbf{x} \in \mathcal{A}$ and $\boldsymbol{\gamma} \in \mathcal{B}$, where

$$
\begin{aligned}
\mathcal{A} &= \left\{\mathbf{x} : ||\mathbf{x}||_2^2 \leq 2P_{\max}\right\}, \\
\mathcal{B} &= \{\boldsymbol{\gamma} \in \mathbb{R}_+^{K \times 1} : ||\boldsymbol{\gamma}||_1 = 1\}.
\end{aligned} \tag{31}
$$

Then (30) is equivalent to

$$
\text{P5}: \max_{\mathbf{x} \in \mathcal{A}} \min_{\boldsymbol{\gamma} \in \mathcal{B}} \ \Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma}), \tag{32}
$$

where

$$
\begin{aligned}
\Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma}) &= \sum_{k=1}^{K-1} \gamma_k \left((\mathbf{r}_k^{[n]})^T \mathbf{x} + t_k^{[n]}\right), \\
&\quad + \frac{\gamma_K ||\mathbf{h}_K||_2^2 \left(2P_{\max} - ||\mathbf{x}||_2^2\right)}{\sigma_K^2}, \\
\mathbf{r}_k^{[n]} &= \frac{2}{\sigma_k^2} \begin{bmatrix} \operatorname{Re}\left(\mathbf{U}^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{U}\mathbf{z}^{[n]}\right) \\ \operatorname{Im}\left(\mathbf{U}^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{U}\mathbf{z}^{[n]}\right) \end{bmatrix}, \quad \forall k \neq K,
\end{aligned}
$$

$$t_k^{[n]} = -\frac{|\mathbf{g}_k^H \mathbf{U} \mathbf{z}^{[n]}|^2}{\sigma_k^2}, \quad \forall k \neq K. \tag{33}$$

Now, it can be seen from P5 that $\Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma})$ is differentiable in both $\mathbf{x}$ and $\boldsymbol{\gamma}$. Moreover, $\Theta^{[n]}$ is a bilevel function of two variables, with the upper layer variable $\mathbf{x}$ constrained by $\ell_2$ norm and the lower layer variable $\boldsymbol{\gamma}$ constrained by $\ell_1$ norm. To address this bilevel problem, the proposed BFOM starts from a feasible $\mathbf{x} = \mathbf{y}^{[0]} \in \mathcal{A}$ and $\boldsymbol{\gamma} = \boldsymbol{\eta}^{[0]} \in \mathcal{B}$ (e.g., $\mathbf{y}^{[0]} = \mathbf{0}$ and $\boldsymbol{\eta}^{[0]} = \mathbf{1}_K/K$), and solves P5 via the following update at the $(m+1)^{\text{th}}$ iteration [44]:

$$\mathbf{y}^\diamond = \underset{\mathbf{x} \in \mathcal{A}}{\operatorname{argmin}} \; W_{\mathcal{A}}\left(\mathbf{x}, \mathbf{y}^{[m]}\right)$$
$$- \frac{1}{2L} \mathbf{x}^T \nabla_{\mathbf{x}} \Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma})\big|_{\mathbf{x} = \mathbf{y}^{[m]}, \boldsymbol{\gamma} = \boldsymbol{\eta}^{[m]}}, \tag{34a}$$

$$\boldsymbol{\eta}^\diamond = \underset{\boldsymbol{\gamma} \in \mathcal{B}}{\operatorname{argmin}} \; W_{\mathcal{B}}\left(\boldsymbol{\gamma}, \boldsymbol{\eta}^{[m]}\right)$$
$$+ \frac{1}{2L} \boldsymbol{\gamma}^T \nabla_{\boldsymbol{\gamma}} \Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma})\big|_{\mathbf{x} = \mathbf{y}^{[m]}, \boldsymbol{\gamma} = \boldsymbol{\eta}^{[m]}}, \tag{34b}$$

$$\mathbf{y}^{[m+1]} = \underset{\mathbf{x} \in \mathcal{A}}{\operatorname{argmin}} \; W_{\mathcal{A}}\left(\mathbf{x}, \mathbf{y}^{[m]}\right)$$
$$- \frac{1}{2L} \mathbf{x}^T \nabla_{\mathbf{x}} \Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma})\big|_{\mathbf{x} = \mathbf{y}^\diamond, \boldsymbol{\gamma} = \boldsymbol{\eta}^\diamond}, \tag{34c}$$

$$\boldsymbol{\eta}^{[m+1]} = \underset{\boldsymbol{\gamma} \in \mathcal{B}}{\operatorname{argmin}} \; W_{\mathcal{B}}\left(\boldsymbol{\gamma}, \boldsymbol{\eta}^{[m]}\right)$$
$$+ \frac{1}{2L} \boldsymbol{\gamma}^T \nabla_{\boldsymbol{\gamma}} \Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma})\big|_{\mathbf{x} = \mathbf{y}^\diamond, \boldsymbol{\gamma} = \boldsymbol{\eta}^\diamond}, \tag{34d}$$

where $W_{\mathcal{A}}$ is the Euclidean distance induced by $\ell_2$ norm and $W_{\mathcal{B}}$ is the KL divergence induced by $\ell_1$ norm:

$$W_{\mathcal{A}}\left(\mathbf{x}, \mathbf{y}^{[m]}\right) = \frac{1}{2}||\mathbf{x} - \mathbf{y}^{[m]}||_2^2, \tag{35}$$

$$W_{\mathcal{B}}\left(\boldsymbol{\gamma}, \boldsymbol{\eta}^{[m]}\right) = \sum_{k=1}^{K} \gamma_k \ln\left(\frac{\gamma_k}{\eta_k^{[m]}}\right). \tag{36}$$

Furthermore, the gradients $\nabla_{\mathbf{x}} \Theta^{[n]}$ and $\nabla_{\boldsymbol{\gamma}} \Theta^{[n]}$ are given by

$$\nabla_{\mathbf{x}} \Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma}) = \sum_{k=1}^{K-1} \gamma_k \mathbf{r}_k^{[n]} - \frac{2\gamma_K ||\mathbf{h}_K||_2^2 \mathbf{x}}{\sigma_K^2}, \tag{37a}$$

$$\nabla_{\boldsymbol{\gamma}} \Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma}) = \Big[(\mathbf{r}_1^{[n]})^T \mathbf{x} + t_1^{[n]}, \cdots, (\mathbf{r}_{K-1}^{[n]})^T \mathbf{x} + t_{K-1}^{[n]},$$
$$||\mathbf{h}_K||_2^2 \left(2P_{\max} - ||\mathbf{x}||_2^2\right) / \sigma_K^2 \Big]^T. \tag{37b}$$

Finally, $L$ is the Bregman Lipschitz constant satisfying

$$||\nabla_{\mathbf{x}} \Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma}) - \nabla_{\mathbf{x}} \Theta^{[n]}(\mathbf{x}', \boldsymbol{\gamma})||_2 \leq L ||\mathbf{x} - \mathbf{x}'||_2, \tag{38a}$$

$$||\nabla_{\mathbf{x}}\Theta^{[n]}\left(\mathbf{x},\boldsymbol{\gamma}\right) - \nabla_{\mathbf{x}}\Theta^{[n]}\left(\mathbf{x},\boldsymbol{\gamma}'\right)||_2 \leq L||\boldsymbol{\gamma} - \boldsymbol{\gamma}'||_1, \tag{38b}$$

$$||\nabla_{\boldsymbol{\gamma}}\Theta^{[n]}\left(\mathbf{x},\boldsymbol{\gamma}\right) - \nabla_{\boldsymbol{\gamma}}\Theta^{[n]}\left(\mathbf{x},\boldsymbol{\gamma}'\right)||_\infty \leq L||\boldsymbol{\gamma} - \boldsymbol{\gamma}'||_1, \tag{38c}$$

$$||\nabla_{\boldsymbol{\gamma}}\Theta^{[n]}\left(\mathbf{x},\boldsymbol{\gamma}\right) - \nabla_{\boldsymbol{\gamma}}\Theta^{[n]}\left(\mathbf{x}',\boldsymbol{\gamma}\right)||_\infty \leq L||\mathbf{x} - \mathbf{x}'||_2, \tag{38d}$$

$$\forall \mathbf{x}, \mathbf{x}' \in \mathcal{A}, \quad \forall \boldsymbol{\gamma}, \boldsymbol{\gamma}' \in \mathcal{B}.$$

The following proposition derives a valid $L$ satisfying the above conditions.

**Proposition 4.** *The conditions* (38a)–(38d) *are satisfied with* $L = \widehat{L}$, *where*

$$\widehat{L} = \max \left[ ||\mathbf{r}_1^{[n]}||_2, \cdots, ||\mathbf{r}_{K-1}^{[n]}||_2, \right.$$
$$\left. 2||\mathbf{h}_K||_2^2 \max\left(\sqrt{2P_{\max}}, 1\right) / \sigma_K^2 \right]. \tag{39}$$

*Proof.* See Appendix E. ■

Although $\widehat{L}$ is a valid Lipschiz constant according to **Proposition 4**, whether a smaller $L$ with $L < \widehat{L}$ would satisfy (38a)–(38d) is not known. Thus, in practice we need to fine-tune the hyper-parameter $L$ and this paper sets $L = \widehat{L}/K$.

The key idea of (34a)–(34d) is to update the variables along their gradient direction, while keeping the updated point $\{\mathbf{y}^\diamond, \boldsymbol{\eta}^\diamond, \mathbf{y}^{[m+1]}, \boldsymbol{\eta}^{[m+1]}\}$ close to the current point $\{\mathbf{y}^{[m]}, \boldsymbol{\eta}^{[m]}\}$. Furthermore, by using the intermediate point $\{\mathbf{y}^\diamond, \boldsymbol{\eta}^\diamond\}$ in (34a)–(34b), we can compute the look-ahead gradient for updating $\{\mathbf{y}^{[m+1]}, \boldsymbol{\eta}^{[m+1]}\}$ as in (34c)–(34d). As proved in [44]–[47], the iterative procedure (34a)–(34d) is guaranteed to converge to the optimal solution to P5 with a convergence rate of $\mathcal{O}(1/m)$.

Notice that the KKT optimality conditions can be adopted to derive the closed-form expressions for (34a)–(34d). In particular, equations (34a)–(34b) are equivalent to

$$\mathbf{y}^\diamond = \left( \max \left[ \left\| \mathbf{y}^{[m]} + \frac{\nabla_{\mathbf{x}}\Theta^{[n]}\left(\mathbf{y}^{[m]}, \boldsymbol{\eta}^{[m]}\right)}{2L} \right\|_2, \sqrt{2P_{\max}} \right] \right)^{-1}$$
$$\times \sqrt{2P_{\max}} \left[ \mathbf{y}^{[m]} + \frac{\nabla_{\mathbf{x}}\Theta^{[n]}\left(\mathbf{y}^{[m]}, \boldsymbol{\eta}^{[m]}\right)}{2L} \right],$$

$$\eta_k^\diamond = \left( \sum_{i=1}^{K} \eta_i^{[m]} \exp\left[ -\frac{\nabla_{\gamma_i}\Theta^{[n]}\left(\mathbf{y}^{[m]}, \boldsymbol{\eta}^{[m]}\right)}{2L} \right] \right)^{-1}$$
$$\times \eta_k^{[m]} \exp\left[ -\frac{\nabla_{\gamma_k}\Theta^{[n]}\left(\mathbf{y}^{[m]}, \boldsymbol{\eta}^{[m]}\right)}{2L} \right], \quad \forall k.$$

On the other hand, the closed-form expressions for (34c)–(34d) can be derived similarly. In terms of computational complexity, the total complexity of the BFOM based AAUC would be $\mathcal{O}(\mathcal{I}MKN)$, where $\mathcal{I}$ and $M$ are the number of iterations for the SCO and BFOM algorithms, respectively.

*Remark 2:* Since our aim is to obtain an approximate solution, we can terminate the iterative procedure when the norm $||\mathbf{y}^{[m+1]} - \mathbf{y}^{[m]}||$ is small enough, e.g., $||\mathbf{y}^{[m]} - \mathbf{y}^{[m-1]}|| < 10^{-4}$, or the number of iterations reaches $3000$. Then, the optimal $\mathbf{x}^*$ to P5 is given by $\mathbf{x}^* = \sqrt{\tau}\,\mathbf{y}^{[m]}/||\mathbf{y}^{[m]}||_2$, where $\tau$ solves

$$\min_{k=1,\cdots,K-1} \frac{\sqrt{\tau}(\mathbf{r}_k^{[n]})^T \mathbf{y}^{[m]}}{||\mathbf{y}^{[m]}||_2} + t_k^{[n]} = \frac{||\mathbf{h}_K||_2^2 \,(2P_{\max} - \tau)}{\sigma_K^2}. \tag{41}$$

Accordingly, the optimal $\mathbf{z}^*$ and $p^*$ to problem (30) can be recovered as $\mathbf{z}^* = [x_1^*, \cdots, x_{N-1}^*]^T + \mathrm{j}\,[x_N^*, \cdots, x_{2N-2}^*]^T$ and $p^* = 2P_{\max} - ||\mathbf{z}^*||_2^2$.

## C. Large Number of Antennas and Users

When both $N$ and $K$ are very large (with $N/K$ being a large constant), we can combine the derivations in Sections V-A and V-B. More specifically, based on $|\mathbf{g}_j^H \mathbf{g}_k|^2/||\mathbf{g}_k||_2^2 \to 0$ for $j \neq k$ when $N \to \infty$ and $\Upsilon(\mathbf{w}) \to \frac{1}{2}\log_2\left(1 + |\mathbf{h}_K^H\mathbf{w}|^2/\sigma_K^2\right)$ when $K \to \infty$, the problem P2 under $N, K \to \infty$ is

$$\text{P6}: \max_{p} \min \left[ \frac{1}{2}\log_2\left(1 + \frac{2P_{\max} - p}{\sum_{i=1}^{K-1}\sigma_i^2/||\mathbf{U}^H\mathbf{g}_i||_2^2}\right), \right.$$
$$\left. \frac{1}{2}\log_2\left(1 + \frac{p||\mathbf{h}_K||_2^2}{\sigma_K^2}\right) \right],$$

where $p$ is the users' total power to be determined. In the objective function, the first term inside the minimum function is monotonic decreasing in $p$ while the second term is monotonic increasing in $p$. Therefore, the optimal $p^*$ to P6 must satisfy

$$\frac{1}{2}\log_2\left(1 + \frac{(2P_{\max} - p)}{\sum_{i=1}^{K-1}\sigma_i^2/||\mathbf{U}^H\mathbf{g}_i||_2^2}\right)$$
$$= \frac{1}{2}\log_2\left(1 + \frac{p||\mathbf{h}_K||_2^2}{\sigma_K^2}\right), \tag{42}$$

which gives

$$p^* = \frac{2P_{\max}}{1 + \frac{||\mathbf{h}_K||_2^2}{\sigma_K^2}\left(\sum_{i=1}^{K-1}\sigma_i^2/||\mathbf{U}^H\mathbf{g}_i||_2^2\right)}. \tag{43}$$

As a result, the optimal $\mathbf{w}^*$ to P2 when $K$ and $N$ go to infinity can be recovered as $\mathbf{w}^* = \sqrt{p^*}\mathbf{h}_K/||\mathbf{h}_K||_2$. With $\mathbf{w}^*$, the optimal $\mathbf{z}^*$ can be computed using **Proposition 2**.

# VI. IRS BASED SCHEME

It can be seen from problem P1 that the proposed AAUC scheme requires extra power consumption from other devices. To avoid extra power consumption, the proposed scheme can be integrated with the emerging technology of intelligent reflecting surfaces (IRS) [48]–[51]. In particular, we can place the IRS elements at the users $k = 1, \cdots, K-1$ (for fair comparison with the non-IRS case). Then each element of the IRS receives the multi-path signals from the BS, and scatters the combined signal with adjustable amplitude and phase to user $K$. By employing the IRS technique, the two transmission phases can be combined into one, and the system cost of the proposed scheme can be significantly reduced.

More specifically, the optimization problem of the IRS based scheme is given by

$$
\text{Q1}: \max_{\mathbf{v},\mathbf{u},R} \quad \frac{1}{D_K} \int_0^{D_K} \mathbb{E}_{\{\phi_k,\delta_k\}} \left\{ \left[ R - \log_2 \left( 1 + \frac{1}{\sigma_E^2} \right. \right.\right.
$$
$$
\left.\left.\left. \left| \mathbf{h}_E^H(\rho, \{\phi_k,\delta_k\}) \left( [\mathbf{g}_1, \cdots, \mathbf{g}_{K-1}]^H \mathbf{v} \circ \mathbf{u} \right) \right|^2 \right) \right]^+ \right\} \mathrm{d}\rho,
$$

$$
\text{s.t.} \quad ||\mathbf{v}||_2^2 \leq P_{\max}, \quad |\mathbf{g}_K^H \mathbf{v}| = 0, \tag{44a}
$$

$$
R \leq \log_2 \left( 1 + \frac{|\mathbf{g}_k^H \mathbf{v}|^2}{\sigma_k^2} \right), \quad \forall k \neq K, \tag{44b}
$$

$$
R \leq \log_2 \left( 1 + \frac{\left| \mathbf{h}_K^H \left( [\mathbf{g}_1, \cdots, \mathbf{g}_{K-1}]^H \mathbf{v} \circ \mathbf{u} \right) \right|^2}{\sigma_K^2} \right), \tag{44c}
$$

$$
|u_k|_2 \leq 1, \quad \forall k \neq K, \tag{44d}
$$

where $\mathbf{u} \in \mathbb{C}^{(K-1)\times 1}$ is the amplitude and phase design at IRS. By introducing a slack variable $\mathbf{c} = [\mathbf{g}_1, \cdots, \mathbf{g}_{K-1}]^H \mathbf{v} \circ \mathbf{u} \in \mathbb{C}^{(K-1)\times 1}$ with $|c_k|^2 \leq |\mathbf{g}_k^H \mathbf{v}|^2$, and adopting the angular secrecy model in (13), the problem Q1 is converted into

$$
\text{Q2}: \max_{\mathbf{v},\mathbf{c},R} \quad \left[ R - \log_2 \left( 1 + \frac{\lambda \mathbf{c}^H \mathbf{J} \mathbf{c}}{\sigma_E^2} \right) \right]^+,
$$

$$
\text{s.t.} \quad ||\mathbf{v}||_2^2 \leq P_{\max}, \quad |\mathbf{g}_K^H \mathbf{v}| = 0, \tag{45a}
$$

$$
R \leq \log_2 \left( 1 + \frac{|\mathbf{g}_k^H \mathbf{v}|^2}{\sigma_k^2} \right), \quad \forall k \neq K, \tag{45b}
$$

$$R \leq \log_2 \left( 1 + \frac{|\mathbf{h}_K^H \mathbf{c}|^2}{\sigma_K^2} \right), \tag{45c}$$

$$|c_k|^2 \leq |\mathbf{g}_k^H \mathbf{v}|^2, \quad \forall k \neq K, \tag{45d}$$

The above problem has a similar structure as P1 and can be solved by the SCO algorithm. Notice that the channel estimation of the IRS based scheme is challenging due to the low costs of IRS elements. Therefore, IRS is not applicable to the case of fast changing channels.

## VII. SIMULATION RESULTS AND DISCUSSIONS

This section provides simulation results to evaluate the performance of the proposed scheme. It is assumed that the noise powers $\sigma_1^2 = \cdots = \sigma_K^2 = \sigma_E^2 = -80 \, \text{dBm}$ [48] (corresponding to power spectral density $-140 \, \text{dBm/Hz}$ [52] with $1 \, \text{MHz}$ bandwidth). The transmit power budget is $P_{\max} = 30 \, \text{dBm}$ and the pathloss exponent is set to $\alpha = 2.5$. The pathloss at the distance of $1 \, \text{m}$ is $-40 \, \text{dB}$, which is computed using the 3GPP UMi model [53] and $3.5 \, \text{GHz}$ carrier frequency (i.e., carrier frequency of 5G in China). We adopt $\lambda = 0.64$ for the secrecy model in (13), and the number of iterations for SCO is set to $\mathcal{I} = 20$.

Based on the above settings, we simulate the following massive MIMO system:

- The BS is located at $(0, 0)$.
- The users are located at $(D_k \cos \theta_k, D_k \sin \theta_k)$ with $D_k \sim \mathcal{U}(100, 500)$ and $\theta_k \sim \mathcal{U}(-\pi, \pi)$ for all $k$.
- The eavesdropper is located at $(D_E \cos \theta_K, D_E \sin \theta_K)$ with $D_E \sim \mathcal{U}(0, D_K)$.
- All the distances use the unit of m.

With $\{D_k, \theta_k, D_E, \theta_E\}$, channels $\{\mathbf{g}_k, \mathbf{g}_E\}$ are generated according to (1). The channel $\mathbf{h}_E$ is generated according to (9), where the distance $d_{E,k}$ between the eavesdropper and the user $k$ is computed based on their locations. The channel $\mathbf{h}_K$ is also generated according to (9), but with $d_{E,k}$ replaced by the distance $d_{K,k}$ between the users $K$ and $k$. Finally, each point in the figures is obtained by averaging over 50 simulation runs, with independent channel realizations and locations of users in each run.

### A. Evaluation of SCO Based AAUC

In order to verify the performance of the proposed SCO based AAUC, the case of $K = 10$ with $N = 100$ is simulated, and the average secrecy rate versus Rician K-factor $K_R \in \{0, 10, 20, 40\}$ in dB is shown in Fig. 3a. Besides the proposed AAUC, we also simulate the direct transmission
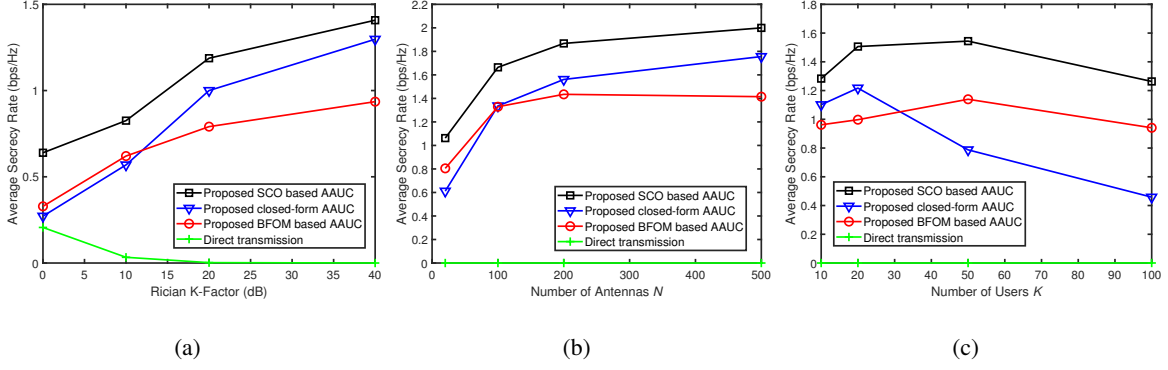
(a)                                    (b)                                    (c)

Fig. 3.  a) Average secrecy rate versus Rician K-factor $K_R$ when $K = 10$ and $N = 100$; b) Average secrecy rate versus the number of antennas $N$ when $K = 20$ and $K_R = 30$ dB; c) Average secrecy rate versus the number of users $K$ when $N = 100$ and $K_R = 30$ dB.



(a)                                    (b)                                    (c)
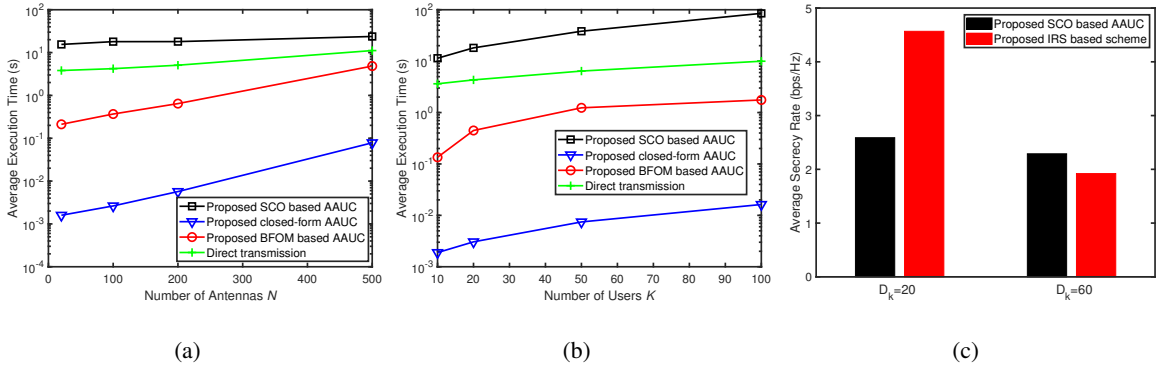
Fig. 4.  a) Average execution time versus the number of antennas $N$; b) Average execution time versus the number of users $K$; c) Average secrecy rate versus the distance $D_k$ when $N = 100$ and $K = 10$.



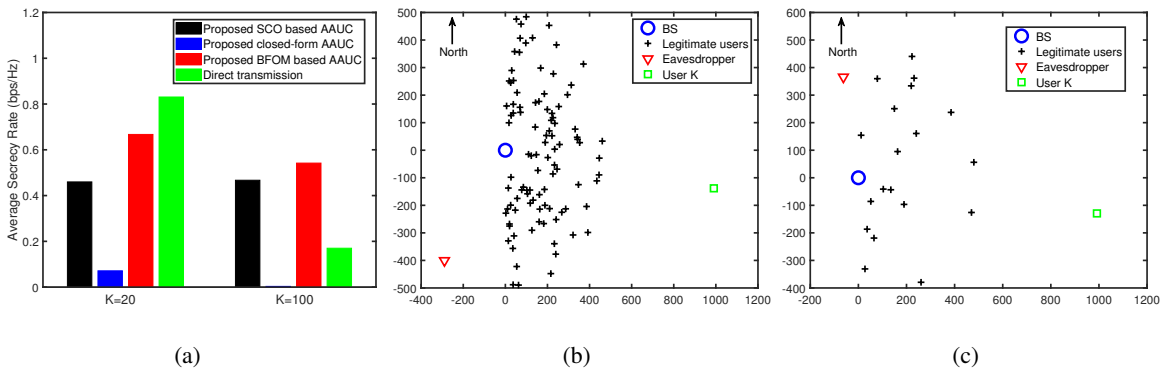(a)                                    (b)                                    (c)

Fig. 5.  a) Average secrecy rate in Rayleigh fading channels with eavesdropper not in the same angle as one of the users; b) A realization of locations of BS, users, and eavesdropper when $K = 100$; c) A realization of locations of BS, users, and eavesdropper when $K = 20$.

scheme. For direct transmission, the multicasting beamforming design $\mathbf{v}$ is obtained by the conic quadratic programming approach [54]. It can be seen from Fig. 3a that when $K_R$ is very small, it is still possible to achieve a positive secrecy rate for the direct transmission scheme, since the channels $\{\mathbf{g}_K, \mathbf{g}_E\}$ are not exactly in the same direction due to the non-LOS components. However, as $K_R$ increases, the direct transmission scheme would lead to zero secrecy rates, since the channels $\mathbf{g}_K$ and $\mathbf{g}_E$ are highly correlated. This result corroborates our discussions in Section I that the direct transmission is vulnerable to the eavesdropping in Rician fading environment. Fortunately, by adopting user cooperation, all the AAUC schemes significantly outperform the direct transmission scheme. Furthermore, no matter what value $K_R$ takes, the proposed SCO based AAUC always outperforms the closed-form AAUC and the BFOM based AAUC. This is because SCO based AAUC does not rely on the assumption of large number of antennas or users. Notice that the secrecy rate of SCO based AAUC increases as $K_R$ increases. This is because $|\mathbf{g}_K^H \mathbf{v}| = 0$ is adopted in P1, and a higher correlation between $\mathbf{g}_K$ and $\mathbf{g}_E$ means a smaller $|\mathbf{g}_E^H \mathbf{v}|^2$ at the eavesdropper.

## B. Evaluation of Large-Scale Optimization Algorithms

Next, to verify the performance and the low complexity nature of closed-form AAUC when $N$ is large, the case of $K = 20$ with $K_R = 30\,\mathrm{dB}$ is simulated, and the average secrecy rate versus the number of antennas $N \in \{20, 100, 200, 500\}$ is shown in Fig. 3b. It can be seen that when $N$ is small (e.g., $N = 20$), the performance gap between SCO based AAUC and closed-form AAUC is large. This is because the channels $\{\mathbf{g}_k\}$ are not orthogonal for a small $N$, which deviates from the orthogonality assumption adopted in the closed-form AAUC. However, as $N$ increases, the performance of closed-form AAUC is approaching that of SCO based AAUC. This corroborates **Proposition 2** and **Proposition 3** derived in Section V-A. On the other hand, Fig. 4a shows the average execution time versus the number of antennas at the BS. Compared to SCO based AAUC, closed-form AAUC saves the computation times by several orders of magnitude for all the simulated values of $N$, revealing the low-complexity nature of the proposed closed-form solution.

To analyze the performance and the complexity of BFOM based AAUC when $K$ is large, the case of $N = 100$ with $K_R = 30\,\mathrm{dB}$ is simulated, and the average secrecy rate versus the number of users $K \in \{10, 20, 50, 100\}$ is shown in Fig. 3c. It can be seen that closed-form AAUC performs poorly in this setting since a large $K$ would lead to nonorthogonal channels of $\{\mathbf{g}_k\}$. In

contrast, BFOM based AAUC achieves a satisfactory secrecy rate. Moreover, the performance gap between BFOM based AAUC and SCO based AAUC decreases as $K$ increases, which corroborates the discussions in Section V-B. On the other hand, as shown in Fig. 4b, BFOM based AAUC saves at least $90\%$ (one order of magnitude) of the computation time compared with SCO based AAUC. Its flat curve of execution time reveals the linear complexity nature of the proposed BFOM algorithm.

### C. Evaluation of IRS Based Scheme

Now, to evaluate the performance of the proposed IRS based scheme, we simulate the case of $N = 100$ and $K = 10$ with $K_R = 70\,\mathrm{dB}$. The average secrecy rate versus the distance $D_k = D_1 = D_2 = \cdots = D_{10}$ is shown in Fig. 4c. It can be seen from Fig. 4c that the proposed IRS based scheme outperforms the proposed SCO based AAUC when $D_k = 20$ due to the smaller number of transmission phases with IRS. However, the IRS based scheme performs worse than the proposed SCO based AAUC when $D_k = 60$. This is because the power of the scattered signal from IRS cannot exceed the received signal power at IRS as seen from (44d). As a result, when the IRS elements are far from the BS, the scattered signal from IRS is weak. In contrast, the helping users in the SCO based AAUC scheme can use a much larger transmit power than its received signal power.

### D. Rayleigh Fading Channel

Finally, to demonstrate the versatility of the proposed method, we simulate a case of non-Rician fading channel and the eavesdropper is not in the same angle as one of the users. In particular, we consider the case of $N = 100$ and $K_R = 0$ (i.e., Rayleigh fading). The eavesdropper is randomly located at $(D_E \cos\theta_E, D_E \sin\theta_E)$ with $D_E \sim \mathcal{U}(100, 500)$ and $\theta_E \sim \mathcal{U}(-\pi, 0)$. The users are randomly located at $(D_k \cos\theta_k, D_k \sin\theta_k)$ with $\theta_k \sim \mathcal{U}(0, \pi)$, where $D_k \sim \mathcal{U}(100, 500)$ for $k \neq K$ and $D_K = 1000$.

In this setting, the proposed AAUC scheme still designs $\mathbf{v}$ and $\mathbf{w}$ by solving P1 (i.e, uses two transmission phases, forces $|\mathbf{g}_K^H \mathbf{v}| = 0$, and adopts the angular secrecy model). The average secrecy rate versus the number of users $K \in \{20, 100\}$ is shown in Fig. 5a. It can be seen that the performance of the closed-form AAUC is not acceptable due to the small $N/K$. On the other hand, the secrecy rates of SCO based AAUC and BFOM based AAUC are significantly higher than that of direct transmission when $K = 100$. This is because the proposed AAUC exploits

path diversity, and the helping users can relay the information to the far-away user as shown in Fig. 5b. However, the proposed AAUC schemes perform slightly worse than traditional direct transmission when $K = 20$ due to $K < N$ (as shown in Fig. 5c). Fortunately, the proposed method can be executed in combination with direct transmission, and the hybrid strategy can select between executing direct transmission or AAUC. By switching to the other transmission mode once the current mode is not satisfactory, the hybrid transmission strategy can always achieve high security.

## VIII. CONCLUSIONS

This paper studied the physical-layer security problem in a massive MIMO multicasting system. Since the objective function in this problem is not analytical, an angular secrecy model, which matches the numerical integration very well, was proposed. By adopting this model, secure beamforming design was obtained via the SCO based AUCC algorithm. In the large-scale settings, two fast algorithms were derived to tackle the curse of high dimensionality. Simulation results showed that the proposed SCO algorithm achieves higher security than direct transmission. Furthermore, the proposed fast algorithms significantly reduce the execution time compared to the SCO while still achieving satisfactory performance.

## APPENDIX A
### PROOF OF PROPERTY (III)

To prove this property, we first notice that

$$
S(R, \mathbf{w}|\sigma_E)
$$

$$
= \frac{1}{D_K} \int_0^{D_K} \mathbb{E}_{\{\phi_k, \delta_k\}} \left\{ \left[ R \right. \right.
$$

$$
\left. \left. - \frac{1}{2}\log_2 \left( 1 + \frac{\left| \mathbf{h}_E^H(\rho, \{\phi_k, \delta_k\})\mathbf{w} \right|^2}{\sigma_E^2} \right) \right]^+ \right\} \mathrm{d}\rho
$$

$$
\geq \left[ R - \frac{1}{D_K} \int_0^{D_K} \mathbb{E}_{\{\phi_k, \delta_k\}} \left\{ \right. \right.
$$

$$
\left. \left. \frac{1}{2}\log_2 \left( 1 + \frac{\left| \mathbf{h}_E^H(\rho, \{\phi_k, \delta_k\})\mathbf{w} \right|^2}{\sigma_E^2} \right) \right\} \mathrm{d}\rho \right]^+
$$

$$\geq \left[ R - \frac{1}{2}\log_2\left( 1 + \frac{1}{\sigma_E^2 D_K} \right. \right.$$

$$\left. \left. \times \int_0^{D_K} \mathbb{E}_{\{\phi_k,\delta_k\}}\left[ \left| \mathbf{h}_E^H(\rho, \{\phi_k, \delta_k\})\mathbf{w} \right|^2 \right] \mathrm{d}\rho \right) \right]^+, \tag{46}$$

where the first inequality is due to the convexity of $[x]^+ = \max(x, 0)$, and the second inequality is due to Jensen's inequality. On the other hand, we compute

$$\frac{1}{D_K} \int_0^{D_K} \mathbb{E}_{\{\phi_k,\delta_k\}}\left[ \left| \mathbf{h}_E^H(\rho, \{\phi_k, \delta_k\})\mathbf{w} \right|^2 \right] \mathrm{d}\rho$$

$$= \frac{1}{D_K} \int_0^{D_K} \mathbb{E}_{\{\phi_k,\delta_k\}}\left[ \left| \sum_{k=1}^{K-1} w_k \sqrt{\varrho_0 \left( \frac{d_{E,k}(\rho)}{d_0} \right)^{-\alpha}} \right. \right.$$

$$\left. \left. \times \left( \sqrt{\frac{K_R}{1+K_R}}\, \mathrm{e}^{\mathrm{j}\phi_k} + \sqrt{\frac{1}{1+K_R}}\, \delta_k \right) \right|^2 \right] \mathrm{d}\rho$$

$$= \frac{1}{D_K} \int_0^{D_K} \left[ \sum_{k=1}^{K-1} |w_k|^2\, \varrho_0 \left( \frac{d_{E,k}(\rho)}{d_0} \right)^{-\alpha} \right] \mathrm{d}\rho$$

$$= \sum_{k=1}^{K-1} |w_k|^2\, \frac{\varrho_0}{D_K} \int_0^{D_K} \left[ (\rho\cos\theta_K - D_k\cos\theta_k)^2 \right.$$

$$\left. + (\rho\sin\theta_K - D_k\sin\theta_k)^2 \right]^{-\alpha/2} \mathrm{d}\rho$$

$$= \mathbf{w}^H \mathbf{J} \mathbf{w}, \tag{47}$$

where the first equality is obtained by putting $\mathbf{h}_E(\rho, \{\phi_k, \delta_k\})$ of (9) into $|\mathbf{h}_E^H(\rho, \{\phi_k, \delta_k\})\mathbf{w}|^2$, and the second equality is due to the independence between $\{\phi_k\}$ and $\{\delta_k\}$ together with $\mathbb{E}[|\delta_k|^2] = 1$, and the third equality is obtained from the expression (8) of $d_{E,k}(\rho)$. Combining (46) and (47), the proof is immediately completed.

## APPENDIX B

### PROOF OF PROPOSITION 1

To prove part (i), consider the following inequality

$$\left( \mathbf{z} - \mathbf{z}^\star \right)^H \frac{\mathbf{U}^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{U}}{\sigma_k^2} \left( \mathbf{z} - \mathbf{z}^\star \right) \geq 0. \tag{48}$$

This always holds due to $\mathbf{U}^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{U} \succeq \mathbf{0}$. Then from (48) we further have

$$\frac{|\mathbf{g}_k^H \mathbf{U} \mathbf{z}|^2}{\sigma_k^2} \geq 2\mathrm{Re}\left[ \frac{(\mathbf{z}^\star)^H \mathbf{U}^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{U} \mathbf{z}}{\sigma_k^2} \right] - \frac{|\mathbf{g}_k^H \mathbf{U} \mathbf{z}^\star|^2}{\sigma_k^2}, \tag{49}$$

which leads to

$$\frac{1}{2}\log_2\left(1+\frac{|\mathbf{g}_k^H\mathbf{U}\mathbf{z}|^2}{\sigma_k^2}\right)$$

$$\geq \frac{1}{2}\log_2\left(1+2\mathrm{Re}\left[\frac{(\mathbf{z}^\star)^H\mathbf{U}^H\mathbf{g}_k\mathbf{g}_k^H\mathbf{U}\mathbf{z}}{\sigma_k^2}\right]-\frac{|\mathbf{g}_k^H\mathbf{U}\mathbf{z}^\star|^2}{\sigma_k^2}\right). \tag{50}$$

On the other hand, since $-1/2\log_2(\cdot)$ is convex, it must be greater than its first-order Taylor expansion, leading to

$$-\frac{1}{2}\log_2\left(1+\frac{\lambda\mathbf{w}^H\mathbf{J}\mathbf{w}}{\sigma_E^2}\right)\geq-\frac{\lambda\mathbf{w}^H\mathbf{J}\mathbf{w}-\lambda(\mathbf{w}^\star)^H\mathbf{J}\mathbf{w}^\star}{2\ln2\left[\sigma_E^2+\lambda(\mathbf{w}^\star)^H\mathbf{J}\mathbf{w}^\star\right]}$$

$$-\frac{1}{2}\log_2\left(1+\frac{\lambda(\mathbf{w}^\star)^H\mathbf{J}\mathbf{w}^\star}{\sigma_E^2}\right). \tag{51}$$

By adding the equations (50) and (51), the inequality $\widetilde{\Phi}_k(\mathbf{z},\mathbf{w}|\mathbf{z}^\star,\mathbf{w}^\star)\leq\Phi_k(\mathbf{z},\mathbf{w})$ is immediately proved. Finally, due to $\mathbf{h}_K\mathbf{h}_K^H\succeq\mathbf{0}$, it is clear that $(\mathbf{w}-\mathbf{w}^\star)^H\frac{\mathbf{h}_K\mathbf{h}_K^H}{\sigma_K^2}(\mathbf{w}-\mathbf{w}^\star)\geq0$ and thus

$$\frac{|\mathbf{h}_K^H\mathbf{w}|^2}{\sigma_K^2}\geq2\mathrm{Re}\left[\frac{(\mathbf{w}^\star)^H\mathbf{h}_K\mathbf{h}_K^H\mathbf{w}}{\sigma_K^2}\right]-\frac{|\mathbf{h}_K^H\mathbf{w}^\star|^2}{\sigma_K^2}. \tag{52}$$

Combining (52) and (51) gives $\widetilde{\Upsilon}(\mathbf{w}|\mathbf{w}^\star)\leq\Upsilon(\mathbf{w})$.

To prove part (ii), we put $\mathbf{z}=\mathbf{z}^\star$ and $\mathbf{w}=\mathbf{w}^\star$ into $\widetilde{\Phi}_k$ in (21) and $\widetilde{\Upsilon}$ in (22). Then we immediately obtain $\widetilde{\Phi}_k(\mathbf{z}^\star,\mathbf{w}^\star|\mathbf{z}^\star,\mathbf{w}^\star)=\Phi_k(\mathbf{z}^\star,\mathbf{w}^\star)$ and $\widetilde{\Upsilon}(\mathbf{w}^\star|\mathbf{w}^\star)=\Upsilon(\mathbf{w}^\star)$.

To prove part (iii), we calculate the following derivatives:

$$\nabla_{\mathbf{z}}\widetilde{\Phi}_k(\mathbf{z},\mathbf{w}|\mathbf{z}^\star,\mathbf{w}^\star)=\frac{\mathbf{U}^H\mathbf{g}_k\mathbf{g}_k^H\mathbf{U}\mathbf{z}^\star}{\sigma_k^2\,2\ln2}$$

$$\left(1+2\mathrm{Re}\left[\frac{(\mathbf{z}^\star)^H\mathbf{U}^H\mathbf{g}_k\mathbf{g}_k^H\mathbf{U}\mathbf{z}}{\sigma_k^2}\right]-\frac{|\mathbf{g}_k^H\mathbf{U}\mathbf{z}^\star|^2}{\sigma_k^2}\right)^{-1}, \tag{53a}$$

$$\nabla_{\mathbf{z}}\Phi_k(\mathbf{z},\mathbf{w})=\left(1+\frac{|\mathbf{g}_k^H\mathbf{U}\mathbf{z}|^2}{\sigma_k^2}\right)^{-1}\frac{\mathbf{U}^H\mathbf{g}_k\mathbf{g}_k^H\mathbf{U}\mathbf{z}}{\sigma_k^2\,2\ln2}, \tag{53b}$$

$$\nabla_{\mathbf{w}}\widetilde{\Phi}_k(\mathbf{z},\mathbf{w}|\mathbf{z}^\star,\mathbf{w}^\star)=-\frac{1}{2\ln2}\frac{\lambda\mathbf{J}\mathbf{w}}{\sigma_E^2+\lambda(\mathbf{w}^\star)^H\mathbf{J}\mathbf{w}^\star}, \tag{53c}$$

$$\nabla_{\mathbf{w}}\Phi_k(\mathbf{z},\mathbf{w})=-\frac{1}{2\ln2}\frac{\lambda\mathbf{J}\mathbf{w}}{\sigma_E^2+\lambda\mathbf{w}^H\mathbf{J}\mathbf{w}}, \tag{53d}$$

$$\nabla_{\mathbf{w}}\widetilde{\Upsilon}(\mathbf{w}|\mathbf{w}^\star)=\left(1+2\mathrm{Re}\left[\frac{(\mathbf{w}^\star)^H\mathbf{h}_K\mathbf{h}_K^H\mathbf{w}}{\sigma_K^2}\right]\right.$$

$$\left.-\frac{|\mathbf{h}_K^H\mathbf{w}^\star|^2}{\sigma_K^2}\right)^{-1}\frac{\mathbf{h}_K\mathbf{h}_K^H\mathbf{w}^\star}{\sigma_k^2\,2\ln2}$$

$$-\frac{1}{2\ln 2}\frac{\lambda\mathbf{Jw}}{\sigma_E^2+\lambda\left(\mathbf{w}^\star\right)^H\mathbf{Jw}^\star}, \tag{53e}$$

$$\nabla_{\mathbf{w}}\Upsilon(\mathbf{w})=\left(1+\frac{|\mathbf{h}_K^H\mathbf{w}|^2}{\sigma_K^2}\right)^{-1}\frac{\mathbf{h}_K\mathbf{h}_K^H\mathbf{w}}{\sigma_K^2\,2\ln 2}$$

$$-\frac{1}{2\ln 2}\frac{\lambda\mathbf{Jw}}{\sigma_E^2+\lambda\mathbf{w}^H\mathbf{Jw}}. \tag{53f}$$

Then by putting $\mathbf{z}=\mathbf{z}^\star$ and $\mathbf{w}=\mathbf{w}^\star$ into (53a)–(53f), the proof for part (iii) is completed.

## APPENDIX C

### PROOF OF PROPOSITION 2

To prove this proposition, we consider the problem P2 with fixed $\mathbf{w}=\mathbf{w}^*$. In such a case, the terms $|\mathbf{h}_K^H\mathbf{w}^*|^2$ and $\lambda(\mathbf{w}^*)^H\mathbf{Jw}^*$ in the objective function of P2 are constants, which can be dropped without changing the solution of $\mathbf{z}$. Further removing $\frac{1}{2}\log_2(\cdot)$ in the objective function due to its monotonicity, P2 with fixed $\mathbf{w}=\mathbf{w}^*$ is transformed into

$$\max_{\mathbf{z}}\ \min_{k\neq K}\ \frac{|\mathbf{g}_k^H\mathbf{Uz}|^2}{\sigma_k^2},$$

$$\text{s.t.}\ \ ||\mathbf{z}||_2^2\leq 2P_{\max}-||\mathbf{w}^*||_2^2. \tag{54}$$

To maximize $\mathbf{z}^H\mathbf{U}^H\mathbf{g}_k\mathbf{g}_k^H\mathbf{Uz}$ for $k\neq K$, the optimal $\mathbf{z}^*\in\mathrm{span}\left(\mathbf{U}^H\mathbf{g}_1,\cdots,\mathbf{U}^H\mathbf{g}_{K-1}\right)$. Therefore, without loss of generality, we can set

$$\mathbf{z}=\sum_{k=1}^{K-1}\sqrt{\xi_k}\,\mathrm{e}^{\mathrm{j}\zeta_k}\frac{\mathbf{U}^H\mathbf{g}_k}{||\mathbf{U}^H\mathbf{g}_k||_2}, \tag{55}$$

with $\{\xi_k\geq 0\}$ being real nonnegative coefficients and $\{\zeta_k\}$ being the corresponding phases.

When $N\to\infty$, we have $|\mathbf{g}_j^H\mathbf{g}_k|^2/||\mathbf{g}_k||_2^2\to 0$ for all $k\neq j$. Adding to $\mathbf{UU}^H\to\mathbf{I}$ as $N\to\infty$, the quantity $|\mathbf{g}_j^H\mathbf{UU}^H\mathbf{g}_k|/||\mathbf{U}^H\mathbf{g}_k||_2\to 0$ for all $k\neq j$. Therefore, by putting $\mathbf{z}$ in (55) into problem (54), we obtain $|\mathbf{g}_k^H\mathbf{Uz}|^2=\xi_k||\mathbf{U}^H\mathbf{g}_k||_2^2$ in the objective function and $||\mathbf{z}||_2^2=\sum_{k=1}^{K-1}\xi_k$ in the constraint, meaning that the phases $\{\zeta_k\}$ would not participate in the optimization. To this end, we can set $\zeta_k=0$ for all $k$ in (55). Finally, putting (55) and $\zeta_k=0$ into (54), problem (54) under $N\to\infty$ is equivalently written as

$$\max_{\boldsymbol{\xi}}\ \min_{k\neq K}\ \frac{\xi_k||\mathbf{U}^H\mathbf{g}_k||_2^2}{\sigma_k^2},$$

$$\text{s.t.}\ \ \sum_{k=1}^{K-1}\xi_k\leq 2P_{\max}-||\mathbf{w}^*||_2^2, \tag{56}$$

where $\boldsymbol{\xi} = [\xi_1, \cdots, \xi_{K-1}]^T \in \mathbb{R}_+^{(K-1) \times 1}$. For the above problem, the optimal $\xi_k^*$ must satisfy

$$\frac{\xi_1 ||\mathbf{U}^H \mathbf{g}_1||_2^2}{\sigma_1^2} = \cdots = \frac{\xi_{K-1} ||\mathbf{U}^H \mathbf{g}_{K-1}||_2^2}{\sigma_{K-1}^2}. \tag{57}$$

Otherwise, we can always increase $\xi_j$ with $j = \mathrm{argmin}_{i \neq K} \ \xi_i ||\mathbf{U}^H \mathbf{g}_i||_2^2 / \sigma_i^2$ by a small quantity $\Delta \xi$ and decrease $\xi_j$ with $j = \mathrm{argmax}_{i \neq K} \ \xi_i ||\mathbf{U}^H \mathbf{g}_i||_2^2 / \sigma_i^2$ by $\Delta \xi$, which would make $\sum_{k=1}^{K-1} \xi_k$ unchanged but increase the objective function of (56). Combining (57) and the constraint of (56), the optimal $\xi_k^*$ to (56) is given by

$$\xi_k^* = \frac{\left(2P_{\max} - ||\mathbf{w}^*||_2^2\right) \sigma_k^2}{\left(\sum_{i=1}^{K-1} \sigma_i^2 / ||\mathbf{U}^H \mathbf{g}_i||_2^2\right) ||\mathbf{U}^H \mathbf{g}_k||_2^2}. \tag{58}$$

Putting $\xi_k = \xi_k^*$ into (55), the proposition is obtained.

## APPENDIX D

### PROOF OF PROPOSITION 3

To solve P3, it can be seen that $\Psi(\mathbf{w})$ inside the minimum function is a monotonically decreasing function of $||\mathbf{w}||_2^2$. Its maximum value $\Psi_{\max} > 0$ is obtained at $||\mathbf{w}||_2^2 = 0$, while its minimum value $\Psi_{\min} < 0$ is obtained at $||\mathbf{w}||_2^2 = 2P_{\max}$. On the other hand, $\Upsilon(\mathbf{w})$ inside the minimum function is a monotonically increasing function of $||\mathbf{w}||_2^2$. Its maximum value $\Upsilon_{\max} > 0$ is obtained at $||\mathbf{w}||_2^2 = 2P_{\max}$, while its minimum value $\Upsilon_{\min} = 0$ is obtained at $||\mathbf{w}||_2^2 = 0$.

Based on the above observations, we now prove that the optimal $\mathbf{w}^*$ satisfies $\Psi(\mathbf{w}^*) = \Upsilon(\mathbf{w}^*)$ by contradiction. In particular, assume that $\Psi(\mathbf{w}^*) < \Upsilon(\mathbf{w}^*)$. Then we must have $\Upsilon(\mathbf{w}^*) > 0$ and thus $||\mathbf{w}^*||_2^2 > 0$. As a result, we can always decrease $||\mathbf{w}||_2^2$ to increase $\Psi(\mathbf{w})$ such that $\min[\Psi(\mathbf{w}), \Upsilon(\mathbf{w})]$ is increased. This is in contradiction to $\mathbf{w}^*$ being optimal. On the other hand, if $\Psi(\mathbf{w}^*) > \Upsilon(\mathbf{w}^*)$, we must have $\Psi(\mathbf{w}^*) > 0$ and thus $||\mathbf{w}^*||_2^2 < 2P_{\max}$. Then we can always increase $||\mathbf{w}||_2^2$ to increase $\Upsilon(\mathbf{w})$ such that $\min[\Psi(\mathbf{w}), \Upsilon(\mathbf{w})]$ is increased. Again, contradiction is established.

Using $\Psi(\mathbf{w}^*) = \Upsilon(\mathbf{w}^*)$, we can add a constraint $\Psi(\mathbf{w}) = \Upsilon(\mathbf{w})$ to P3 without changing the problem, and P3 is equivalently written as

$$\max_{\mathbf{w}} \ \Upsilon(\mathbf{w}), \quad \text{s.t.} \ \Psi(\mathbf{w}) = \Upsilon(\mathbf{w}). \tag{59}$$

By dropping the logarithm function in $\Upsilon$ and rearranging $\Psi(\mathbf{w}) = \Upsilon(\mathbf{w})$, the above problem is further simplified into

$$\max_{\mathbf{w}} \ \frac{1 + |\mathbf{h}_K^H \mathbf{w}|^2 / \sigma_K^2}{1 + \lambda \mathbf{w}^H \mathbf{J} \mathbf{w} / \sigma_E^2}, \quad \text{s.t.} \ \mathbf{w}^H \boldsymbol{\Xi} \mathbf{w} = 1, \tag{60}$$

where the constraint is obtained from $\Psi(\mathbf{w}) = \Upsilon(\mathbf{w})$ and $\Xi$ is given by (28). Using $\mathbf{w}^H \Xi \mathbf{w} = 1$, the objective function of (60) is equal to

$$\mathbf{w}^H(\Xi + \mathbf{h}_K \mathbf{h}_K^H / \sigma_K^2)\mathbf{w} \Big/ \left[\mathbf{w}^H(\Xi + \lambda \mathbf{J}/\sigma_E^2)\mathbf{w}\right].$$

Furthermore, by dropping the constraint $\mathbf{w}^H \Xi \mathbf{w} = 1$, the problem (60) is relaxed into

$$\max_{\mathbf{w}} \frac{\mathbf{w}^H(\Xi + \mathbf{h}_K \mathbf{h}_K^H / \sigma_K^2)\mathbf{w}}{\mathbf{w}^H(\Xi + \lambda \mathbf{J}/\sigma_E^2)\mathbf{w}}. \tag{61}$$

It can be seen that the objective function of (61) would not change if we scale $\mathbf{w}$ by any positive factor. As a result, no matter what value the optimal solution of $\mathbf{w}$ to (61) takes, we can always scale it such that $\mathbf{w}^H \Xi \mathbf{w} = 1$ is satisfied. This means that the relaxed problem (61) is equivalent to (60), and we can focus on solving (61) in the following.

To solve (61), a new variable $\mathbf{q} = (\Xi + \lambda \mathbf{J}/\sigma_E^2)^{1/2}\mathbf{w}$ is introduced to replace $\mathbf{w}$. Then (61) is transformed into

$$\max_{\mathbf{q}} \mathbf{q}^H(\Xi + \lambda \mathbf{J}/\sigma_E^2)^{-1/2}(\Xi + \mathbf{h}_K \mathbf{h}_K^H / \sigma_K^2)$$

$$\times (\Xi + \lambda \mathbf{J}/\sigma_E^2)^{-1/2}\mathbf{q} \Big/ \mathbf{q}^H \mathbf{q}, \tag{62}$$

which is the standard eigenvalue problem, and the optimal $\mathbf{q}^*$ is the dominant eigenvector of $(\Xi + \lambda \mathbf{J}/\sigma_E^2)^{-1/2}(\Xi + \mathbf{h}_K \mathbf{h}_K^H / \sigma_K^2)(\Xi + \lambda \mathbf{J}/\sigma_E^2)^{-1/2}$. Putting $\mathbf{q} = \mathbf{q}^*$ into $\mathbf{q} = (\Xi + \lambda \mathbf{J}/\sigma_E^2)^{1/2}\mathbf{w}$ and scaling $\mathbf{w}$ such that $\mathbf{w}^H \Xi \mathbf{w} = 1$, the equation (27) is obtained. This completes the proof.

## APPENDIX E

### PROOF OF PROPOSITION 4

To prove this proposition, we first compute

$$||\nabla_{\mathbf{x}}\Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma}) - \nabla_{\mathbf{x}}\Theta^{[n]}(\mathbf{x}', \boldsymbol{\gamma})||_2$$

$$= \frac{2\gamma_K ||\mathbf{h}_K||_2^2}{\sigma_K^2}||\mathbf{x} - \mathbf{x}'||_2 \le \frac{2||\mathbf{h}_K||_2^2}{\sigma_K^2}||\mathbf{x} - \mathbf{x}'||_2. \tag{63}$$

It can be seen from (63) and (39) that $\widehat{L}$ satisfies (38a). Next, we will prove that $\widehat{L}$ satisfies (38b). In particular, based on the gradient in (37a), we have

$$||\nabla_{\mathbf{x}}\Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma}) - \nabla_{\mathbf{x}}\Theta^{[n]}(\mathbf{x}, \boldsymbol{\gamma}')||_2$$

$$= \left|\left| \sum_{k=1}^{K-1}(\gamma_k - \gamma_k')\mathbf{r}_k^{[n]} - (\gamma_K - \gamma_K')\frac{2||\mathbf{h}_K||_2^2\mathbf{x}}{\sigma_K^2} \right|\right|_2$$

$$\leq \sum_{k=1}^{K-1} |\gamma_k - \gamma'_k| \, ||\mathbf{r}_k^{[n]}||_2 + |\gamma_K - \gamma'_K| \frac{2||\mathbf{h}_K||_2^2 ||\mathbf{x}||_2}{\sigma_K^2}$$

$$\leq \sum_{k=1}^{K-1} |\gamma_k - \gamma'_k| \, ||\mathbf{r}_k^{[n]}||_2 + |\gamma_K - \gamma'_K| \frac{2\sqrt{2P_{\max}}||\mathbf{h}_K||_2^2}{\sigma_K^2}$$

$$\leq \widehat{L} \sum_{k=1}^{K} |\gamma_k - \gamma'_k| = \widehat{L}||\boldsymbol{\gamma} - \boldsymbol{\gamma}'||_1, \tag{64}$$

where the first inequality is due to $||\mathbf{a}+\mathbf{b}|| \leq ||\mathbf{a}|| + ||\mathbf{b}||$, the second inequality is due to $\mathbf{x} \in \mathcal{A}$, and the last inequality is due to $||\mathbf{r}_k^{[n]}||_2 \leq \widehat{L}$ and $2\sqrt{2P_{\max}}||\mathbf{h}_K||_2^2/\sigma_K^2 \leq \widehat{L}$.

Finally, it is clear that (38c) holds since $||\nabla_{\boldsymbol{\gamma}}\Theta^{[n]}(\mathbf{x},\boldsymbol{\gamma}) - \nabla_{\boldsymbol{\gamma}}\Theta^{[n]}(\mathbf{x},\boldsymbol{\gamma}')||_\infty = 0$. Therefore, we only need to show that $\widehat{L}$ satisfies (38d). To this end, based on the gradient in (37b), the left hand side of (38d) is equal to

$$||\nabla_{\boldsymbol{\gamma}}\Theta^{[n]}(\mathbf{x},\boldsymbol{\gamma}) - \nabla_{\boldsymbol{\gamma}}\Theta^{[n]}(\mathbf{x}',\boldsymbol{\gamma})||_\infty$$

$$= \max \left[ \left|(\mathbf{r}_1^{[n]})^T(\mathbf{x}-\mathbf{x}')\right|, \cdots, \left|(\mathbf{r}_{K-1}^{[n]})^T(\mathbf{x}-\mathbf{x}')\right|, \right.$$

$$\left. \frac{||\mathbf{h}_K||_2^2}{\sigma_K^2}\left|||\mathbf{x}||_2^2 - ||\mathbf{x}'||_2^2\right| \right]$$

$$\leq \max \left[ ||\mathbf{r}_1^{[n]}||_2 \, ||\mathbf{x}-\mathbf{x}'||_2, \cdots, ||\mathbf{r}_{K-1}^{[n]}||_2 \, ||\mathbf{x}-\mathbf{x}'||_2, \right.$$

$$\left. \frac{||\mathbf{h}_K||_2^2}{\sigma_K^2}\left|||\mathbf{x}||_2^2 - ||\mathbf{x}'||_2^2\right| \right], \tag{65}$$

where the inequality is due to $|\mathbf{a}^T\mathbf{b}| \leq ||\mathbf{a}||_2||\mathbf{b}||_2$. Moreover,

$$\frac{||\mathbf{h}_K||_2^2}{\sigma_K^2}\left|||\mathbf{x}||_2^2 - ||\mathbf{x}'||_2^2\right|$$

$$= \frac{||\mathbf{h}_K||_2^2}{\sigma_K^2}(||\mathbf{x}||_2 + ||\mathbf{x}'||_2)\left|||\mathbf{x}||_2 - ||\mathbf{x}'||_2\right|$$

$$\leq \frac{||\mathbf{h}_K||_2^2}{\sigma_K^2} \times 2\sqrt{2P_{\max}}||\mathbf{x}-\mathbf{x}'||_2, \tag{66}$$

where the inequality is due to $\mathbf{x},\mathbf{x}' \in \mathcal{A}$ and $\left|||\mathbf{a}|| - ||\mathbf{b}||\right| \leq ||\mathbf{a}-\mathbf{b}||$. By putting (66) into (65), the proof is completed.

## REFERENCES

[1] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Multiple antenna technologies for beyond 5G," [Online]. Available: arXiv: 1910.00092, Sep. 2019.

[2] A. Mukherjee, A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1550–1573, 3rd Quarter, 2014.

[3] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *IEEE Network*, vol. 28, no. 4, pp. 46–50, Aug. 2014.

[4] X. Chen, D. W. K. Ng, W. Gerstacker, and H. H. Chen, "A survey on multiple-antenna techniques for physical layer security," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1027–1053, Jun. 2017.

[5] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[6] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, Aug. 2011.

[7] K. Cumanan, Z. Ding, M. Xu, and H. V. Poor, "Secrecy rate optimization for secure multicast communications," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 8, pp. 1417–1432, Dec. 2016.

[8] Z. Li, S. Wang, P. Mu, and Y.-C. Wu, "Probabilistic constrained secure transmission: Variable-rate design and performance analysis," *IEEE Trans. Wireless Commun.*, early access, Jan. 2020. DOI: 10.1109/TWC.2020.2966476.

[9] H. Q. Ngo, H. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sep. 2014.

[10] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[11] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo. "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99878–99888, Jul. 2019.

[12] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 21–27, Jun. 2015.

[13] Y. Wu, A. Khisti, C. Xiao, G. Caire, K.-K. Wong, and X. Gao, "A survey of physical layer security techniques for 5G wireless networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 679–695, Apr. 2018.

[14] X. Chen, Z. Zhang, C. Zhong, D. W. K. Ng, and R. Jia, "Exploiting inter-user interference for secure massive non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 788–801, Apr. 2018.

[15] B. Akgun, M. Krunz, and O. O. Koyluoglu, "Vulnerabilities of massive MIMO systems to pilot contamination attacks," *IEEE Trans. Inf. Forensic Secur.*, vol. 14, no. 5, pp. 1251–1263, May 2019.

[16] Y. O. Basciftci, C. E. Koksal, and A. Ashikhmin, "Physical-layer security in TDD massive MIMO," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7359–7380, Nov. 2018.

[17] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, "Secure massive MIMO transmission with an active eavesdropper," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3880–3890, Jul. 2016.

[18] J. Zhu, R. Schober, and V. K. Bhargava, "Linear precoding of data and artificial noise in secure massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2245–2261, Mar. 2016.

[19] N. P. Nguyen, H. Q. Ngo, T. Q. Duong, H. D. Tuan, and K. Tourki, "Secure massive MIMO with the artificial noise-aided downlink training," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 802–816, Apr. 2018.

[20] Y.-Y. Zhang, J.-K. Zhang, and H.-Y. Yu, "Physically securing energy-based massive MIMO MAC via joint alignment of multi-user constellations and artificial noise," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 829–844, Apr. 2018.

[21] F. Zhu, F. Gao, H. Lin, S. Jin, J. Zhao, and G. Qian, "Robust beamforming for physical layer security in BDMA massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 775–787, Apr. 2018.

[22] J. Zhang, L. Dai, X. Zhang, E. Björnson, and Z. Wang, "Achievable rate of Rician large-scale MIMO channels with transceiver hardware impairments," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8800–8806, Oct. 2016.

[23] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.

[24] H. Yang and T. L. Marzetta, "Massive MIMO with max-min power control in line-of-sight propagation environment," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4685–4693, Nov. 2017.

[25] A. Chaman, J. Wang, J. Sun, H. Hassanieh, and R. R. Choudhury, "Ghostbuster: Detecting the presence of hidden eavesdroppers," in *Proc. ACM Mobicom*, New Delhi, India, 2018, pp. 337–351.

[26] A. A. Khuwaja, Y. Chen, N. Zhao, M. S. Alouini, and P. Dobbins, "A survey of channel modeling for UAV communications," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2804–2821, Jul. 2018.

[27] J. Zhang, L. Dai, Z. He, S. Jin, and X. Li, "Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327–1338, Jun. 2017.

[28] M. Malmirchegini and Y. Mostofi, "On the spatial predictability of communication channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 964–978, Mar. 2012.

[29] E. Karipidis, N. D. Sidiropoulos, and Z. Q. Luo, "Far-field multicast beamforming for uniform linear antenna arrays," *IEEE Trans. Signal Process.*, vol 55, no. 10, pp. 4916–4927, Oct. 2007.

[30] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[31] D. W. K. Ng, E. S. Lo, and R. Schober, "Robust beamforming for secure communication in systems with wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4599–4615, Aug. 2014.

[32] A. Rabbachin, A. Conti, and M. Z. Win, "Wireless network intrinsic secrecy," *IEEE/ACM Trans. Netw.*, vol. 23, no. 1, pp. 56–69, Feb. 2015.

[33] A. Mukherjee, "Physical-layer security in the Internet of Things: Sensing and communication confidentiality under resource constraints," *Proc. IEEE*, vol. 103, no. 10, pp. 1747–1761, Oct. 2015.

[34] L. Mucchi, L. Ronga, X. Zhou, K. Huang, Y. Chen, and R. Wang, "A new metric for measuring the security of an environment: The secrecy pressure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3416–3430, May 2017.

[35] X. Sun, D. W. K. Ng, Z. Ding, Y. Xu, and Z. Zhong, "Physical layer security in UAV systems: Challenges and opportunities," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 40–47 Oct. 2019.

[36] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operation Research*, vol. 26, no. 4, Jul. 1978.

[37] T. Lipp and S. Boyd, "Variations and extensions of the convex-concave procedure," *Optimization and Engineering*, vol. 17, no. 2, pp. 263–287, Jun. 2016.

[38] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76–88, Jan. 2016.

[39] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[40] H. A. Le Thi and P. D. Tao, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Annals of Operations Research*, vol. 133, no. 1–4, pp. 23–46, Jan. 2005.

[41] S. Wang, M. Xia, and Y.-C. Wu, "Multipair two-way relay network with harvest-then-transmit users: resolving pairwise uplink-downlink coupling," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 8, pp. 1506–1521, Dec. 2016.

[42] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[43] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization* (MPS/SIAM Series on Optimizations). Philadelphia, PA, USA: SIAM, 2013.

[44] A. Nemirovski, "Prox-method with rate of convergence $\mathcal{O}(1/T)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems," *SIAM J. Optimiz.*, vol. 15, no. 1, pp. 229–251, 2004.

[45] S. Bubeck, "Convex optimization: Algorithms and complexity," *Foundations and Trends in Machine Learning*, vol. 8, no. 3–4, pp. 231–357, Nov. 2015.

[46] A. Konar and N. D. Sidiropoulos, "Fast approximation algorithms for a class of non-convex QCQP problems using first-order methods," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3494–3509, Jul. 2017.

[47] S. Wang, Y.-C. Wu, M. Xia, R. Wang, and H. V. Poor, "Machine intelligence at the edge with learning centric power allocation," Nov. 2019. [Online]. Available: https://arxiv.org/pdf/1911.04922.pdf.

[48] 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. 2019.

[49] 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, Aug. 2019.

[50] X. Yu, D. Xu, Y, Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," [Online]. Available: arXiv:1912.01497, Dec. 2019.

[51] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource allocation for secure IRS-assisted multiuser MISO systems," *IEEE Globecom Workshops*, Dec. 2019. [Online]. Available: arXiv:1907.03085.

[52] S. Wang, M. Xia, K. Huang, and Y.-C. Wu, "Wirelessly powered two-way communication with nonlinear energy harvesting model: Rate regions under fixed and mobile relay," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8190–8204, Dec. 2017.

[53] 3GPP, "Technical specification group radio access network; study on 3d channel model for LTE (release 12)," TR 36.873 V12.7.0, Dec. 2017.

[54] L. N. Tran, M. F. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 114–117, Jan. 2014.