Designing IRS-Aided MIMO Systems for Secrecy Enhancement

Saba Asaad, Member, IEEE, Yifei Wu, Ali Bereyhi Member, IEEE, Ralf R. Müller, Senior Member, IEEE, Rafael F. Schaefer, Senior Member, IEEE, and H. Vincent Poor, Fellow, IEEE

Abstract

Intelligent reflecting surfaces (IRSs) enable multiple-input multiple-output (MIMO) transmitters to modify the communication channels between the transmitters and receivers. In the presence of eavesdropping terminals, this degree of freedom can be used to effectively suppress the information leakage towards such malicious terminals. This leads to significant potential secrecy gains in IRS-aided MIMO systems. This work exploits these gains via a tractable joint design of downlink beamformers and IRS phase-shifts. In this respect, we consider a generic IRS-aided MIMO wiretap setting and invoke fractional programming and alternating optimization techniques to iteratively find the beamformers and phase-shifts that maximize the achievable weighted secrecy sum-rate. Our design concludes two low-complexity algorithms for joint beamforming and phase-shift tuning. Performance of the proposed algorithms are numerically evaluated and compared to the benchmark. The results reveal that integrating IRSs into MIMO systems not only boosts the secrecy performance of the system, but also improves the robustness against passive eavesdropping.

Index Terms

Intelligent reflecting surfaces, physical layer security, fractional programming, block coordinate decent, majorization-maximization method, alternating optimization.

Saba Asaad, Yifei Wu, Ali Bereyhi and Ralf R. Müller are with the Institute for Digital Communications, Friedrich-Alexander Universität Erlangen-Nürnberg, Germany, {saba.asaad, yifei.wu, ali.bereyhi, ralf.r.mueller}@fau.de. Rafael F. Schaefer is with the Chair of Communications Engineering and Security, University of Siegen, Germany, rafael.schaefer@uni-siegen.de. H. Vincent Poor is with the Department of Electrical and Computer Engineering, Princeton University, NJ 08544, USA, poor@princeton.edu.

This work was supported by Deutsche Forschungsgemeinschaft (DFG) under Grant MU 3735/7-1 Project-No. 409561515, and in part by the German Federal Ministry for Education and Research (BMBF) under Grant 16KIS1242.
I. INTRODUCTION

Over the past few years, the unprecedented growth of data traffic due to the popularity of smart devices has created many challenges in the design of next-generation wireless networks [1]–[3]. To address these challenges, various advanced technologies have been proposed in recent years. Examples of such technologies are massive multiple-input multiple-output (MIMO) systems [4]–[6], millimeter wave (mmW) communications [7]–[9], and ultra-dense heterogeneous networks [10]–[12], just to name a few. Although employing these key technologies can significantly enhance the spectral efficiency of wireless networks, practical limits related to energy consumption, hardware cost, and transceiver complexity are still considered burdensome [13], [14]. Intelligent reflecting surface (IRS)-aided MIMO communications has recently attracted considerable attention as an effective solution to these implementational issues [15], [16].

Recent advances in the design of reflecting surfaces have introduced IRSs as efficient components which can help overcoming various challenges of earlier enabling technologies [17]. In fact, the ability of IRSs in beamforming and reflecting the received signals with neither noise amplification nor self-interference make them different from basic relaying components, e.g., conventional repeaters, in wireless networks [18]. IRSs are flexibly installed on room ceiling, interior walls, and building facades [19]. Hence, from the implementational viewpoint, IRS-aided architectures can significantly reduce implementation cost without considerable performance degradation.

A primary application of IRSs in wireless communications is to employ them for realizing a massive MIMO transmitter in a distributed fashion [20], [21]: Instead of gathering a vast number of antennas at a single base station (BS), an IRS is placed out of the location of a BS with a limited number of antennas. Despite conceptual differences to massive MIMO, such settings are shown to sustain the fundamental features of massive MIMO systems while enjoying low implementational complexity [22], [23]. IRSs can further enable cost-efficient realizations of hybrid analog digital architectures [24]–[27]. In such architectures, the analog network is implemented via a set of radio frequency (RF) transmitters illuminating the digitally precoded signals towards an IRS, and the analog beamforming is performed by proper phase-shifting at the IRS. As the physical analog network is replaced via a "noise-free wireless channel," these architectures significantly reduce the power-loss. Interestingly, this gain is obtained at no considerable degradation in the quality of downlink transmission [24].
In addition to the mentioned applications, IRSs potentially provide benefits in terms of secrecy [28]–[30], network coverage [31] and spectrum efficiency [32], [33]. A comprehensive overview of IRS-aided wireless communications is provided in [34] where the authors discuss main applications, channel estimation, deployment challenges and hardware issues.

A. Secrecy in IRS-aided MIMO Communications

In an IRS-aided wireless network, the phase-shifts of the elements at the IRSs can be tuned such that the reflected signals from the IRSs will be

- **constructively** combined with the copy of the transmitted signal received through the direct links between the BS and the legitimate terminals, and
- **destructively** added to the signals received by eavesdroppers via their respective direct links.

This property enables IRSs to greatly enhance the physical layer security of MIMO settings. This potential however comes at the price of design complexity. In fact, to employ effectively an IRS, one needs to jointly perform beamforming and phase-shift tuning. This often lead to a computationally intractable optimization; see for example discussions in [35]. As a result, most primary studies consider the design of IRS-aided MIMO systems for simplified settings; see for instance [35]–[37].

A fundamental study on the secrecy performance of IRS-aided settings is given in [35], where the authors considered a simple wiretap setting with a single legitimate receiver and a single eavesdropper. The investigations are further extended to a slightly different setting with a superior eavesdropping channel in [36]. The study in [38] investigates an IRS-aided multiple-input single-output (MISO) transmission in a basic wiretap channel with a single-antenna eavesdropper. Using deep learning for phase-shift tuning and downlink beamforming is considered in [39], and the performance is compared with the conventional methods in [35] and [36].

In addition to the passive beamforming gain achieved by IRSs, the extra degrees of freedom provided by these surfaces lead to secrecy enhancements in various other respects. For instance, the studies in [30], [40], [41] demonstrate that generating artificial noise and applying controlled jamming by the BS in IRS-assisted MISO settings result in significantly better secrecy performance than in classic settings without IRSs. Another example is the robustness of IRS-assisted MIMO settings against active eavesdropping reported recently in [42]. In this study, an IRS-assisted wiretap setting with active eavesdroppers is considered that contaminate the uplink training pilots of the legitimate user terminals (UTs) to increase their received information
leakage. The results of this study reveal that the further degrees of freedom achieved by the IRS enables us to statistically blind the active eavesdroppers. This is in contrast to classical massive MIMO settings that are known to be non-robust against active eavesdropping; see [43]–[45] and the references therein.

The basic studies on secrecy performance of IRS-aided MIMO settings are further extended in multiple directions; see for instance [46]–[48]. In [46], the authors investigate the secrecy performance of a cooperative jamming strategy applied by multiple eavesdroppers employing an IRS to attack a single-antenna legitimate UT. Secure downlink transmission in a multiuser IRS-aided MISO system with multiple legitimate and malicious UTs is further studied in [47]. Despite extending the basic investigations to multiuser settings, the work considers a special scenario in which all UTs are located in the same direction of the transmitter and hence their channels are highly correlated. The work is extended in [48] with respect to the channel model. The latter study is still restricted to cases with blocked direct channels from the transmitter to the IRSs and eavesdroppers.

B. Contributions

As mentioned, the results in the literature imply that integrating IRSs into MIMO systems can lead to significant enhancements in terms of secrecy performance. Although this finding is concluded from several recent lines of work, a design of secure active and passive beamforming in IRS-aided MIMO systems is yet to be addressed. The main goal of this work is to design a secure joint active, i.e., precoding at the transmitter, and passive beamforming, i.e., phase-shift tuning at IRSs, by which the potential secrecy gains of IRS-aided settings are exploited. This goal is reached through the following contributions:

- First the secrecy performance of an IRS-aided multiuser MIMO wiretap setting is described in terms of the achievable weighted secrecy sum-rate. The beamforming design is then formulated as a non-convex optimization problem. To cope with the non-convexity, we initially derive a parameterized variational problem whose solution returns the optimal design. We then invoke fractional programming (FP) [49], [50] to transform the variational problem into a quadratic problem. FP has been widely used for throughput maximization in communication systems. The most relevant instance is the work in [51], in which the authors use FP for weighted sum-rate maximization in an IRS-aided MIMO system. Although the derivations in [51] consider a generic IRS-aided MIMO system, extending them to a
wiretap setting is not straightforward. We address this issue by dividing the variational optimization into multiple marginal problems and by conditionally bounding the objective of each marginal problem. The final solution is then found by following the alternating optimization (AO) strategy.

- Using AO, the final solution is approximated via an iterative algorithm which deals with multiple inner loops. This leads to high computational complexity. We hence develop two algorithms with reduced complexity; namely a two-tiers and a single-loop algorithm. For each of these algorithms, we show that the weighted secrecy sum-rate increases monotonically as the algorithm iterates.

- To confirm the validity of our derivations, we investigate the performance of the proposed algorithms by running several numerical experiments. The results are compared to multiple reference scenarios, e.g., the scenarios with no IRS and random phase shift, as well as the state-of-the-art. Our investigations confirm that the proposed scheme exploits the potential secrecy gains of IRS-aided systems efficiently, while imposing a tractable processing load on to the system.

C. Notation and Organization

Scalars, vectors and matrices are represented with non-bold, bold lower-case, and bold upper-case letters, respectively. $H^H$ indicates the transposed conjugate of $H$ and $I_N$ is an $N \times N$ identity matrix. The $\ell^2$-norm of $x$ is denoted by $\|x\|$, and $\|W\|_F$ represents the Frobenius norm of matrix $W$. log indicates logarithm to the base 2. $\mathbb{R}$ and $\mathbb{C}$ refer to the real axis and the complex plane, respectively. For $z \in \mathbb{C}$, $z^*$, $\Re\{z\}$, and $\Im\{z\}$ denote the complex conjugate, real part, and imaginary part of $z$, respectively. $\mathcal{CN}(\eta, \sigma^2)$ represents the complex Gaussian distribution with mean $\eta$ and variance $\sigma^2$. For sake of brevity, $\{1, \ldots, N\}$ is shown by $[N]$.

The remaining parts of this manuscript are organized as follows: The system model is described in Section II. In Section III, the design problem is formulated and a variational problem with analytic objective is derived. Section IV gives a quick introduction to FP and the key analytical tools used in the paper. A class of alternative algorithms based on the AO strategy is then derived in Section V. Section VI proposes algorithms with reduced complexity and discusses their convergence. Numerical investigations are presented in Section VII. Finally, the manuscript is concluded in Section VIII.
Secure downlink transmission in an IRS-aided broadcast setting is considered. For sake of brevity, we focus on a single cell in which a BS with $M$ transmit antennas serves $K$ single-antenna legitimate users. $J$ single-antenna eavesdroppers passively overhear the channel. To improve the downlink communication links, a passive IRS unit with $N$ programmable phase shifters is deployed. The elements on this unit receive copies of the signal transmitted by the BS and reflect them after applying phase shifts. These phase shifts are adjustable and controlled by a central control unit. A schematic diagram of the setting is represented in Fig. 1.

We consider transmission over a quasi-static slow fading channel which models either a narrow-band single-carrier system or a particular sub-channel of a wide-band multi-carrier system. The system operates in the time division duplexing (TDD) mode. This means that the uplink and downlink transmissions are performed at the same carrier frequency. As the result, the uplink and downlink channels are reciprocal. Following this property, the BS acquires the channel state information (CSI) in the uplink training phase and directly applies it to the downlink signal transmission.

For channel estimation, we assume that the transmitter uses a classic channel estimation
algorithm for an IRS-aided MIMO system, e.g., [52], [53]. It is further assumed that the eavesdroppers are registered users in the system. This means that during the uplink training phase, the BS also acquires the CSI of the eavesdroppers. To keep the analysis tractable, we further ignore the channel estimation error and accordingly assume that the CSI of the receive terminals are perfectly available at the BS.

Remark 1. It is worth mentioning that in this work we focus on passive eavesdropping. Such a setting models broadcast scenarios with confidential messages in which the eavesdroppers are registered UTs. These UTs are supposed to receive the common messages broadcasted in the network, but should receive no information regarding the confidential messages; see for example studies in [54], [55] and references therein. The act of eavesdropping in such networks is conceptually different from overhearing the common message. In the latter case, eavesdroppers are not registered users and are purely employed to eavesdrop upon a message which is supposed to be received only by the registered users. In such a scenario, the eavesdroppers either do not participate in the uplink training phase\(^1\) or perform an active attack\(^2\), e.g., the active pilot attack; see for instance the system model in [42].

A. System Model

To model the downlink transmission in this setting, we note that each UT receives a superposition of two signals:

1) A signal which is received through the direct path\(^3\) between the BS and the UT, and
2) a signal which is reflected via the IRS.

Let \( \mathbf{x} \in \mathbb{C}^M \) contain the signal samples being transmitted in a particular symbol interval via the transmit antennas at the BS. The received signal at legitimate UT \( k \in [K] \) is hence given by

\[
y_k = \mathbf{h}_{d,k}^H \mathbf{x} + \mathbf{h}_{r,k}^H \mathbf{\Phi}^H \mathbf{r} + \vartheta_k. \tag{1}
\]

In (1), \( \mathbf{\Phi}, \mathbf{r}, \vartheta_k, \mathbf{h}_{d,k}, \) and \( \mathbf{h}_{r,k} \) are defined as follows:

- \( \mathbf{\Phi} \in \mathbb{C}^{N \times N} \) is a diagonal matrix modeling the phase-shifts applied by the elements on the IRS-unit,

\(^1\)Hence, their CSI is not available at the BS.

\(^2\)And hence, they contaminate the acquired CSI.

\(^3\)Note that the direct path is not necessarily the line-of-sight. It merely refers to the channel between the BS and a UT which may include a line-of-sight path and/or scatterings.
• $r$ is the signal received by the IRS-unit from the BS,

• $\vartheta_k$ is complex Gaussian noise with zero mean and variance $\sigma_k^2$, i.e., $\vartheta_k \sim \mathcal{CN}(0, \sigma_k^2)$,

• $h_{d,k} \in \mathbb{C}^M$ represents the conjugate of direct uplink channel between UT $k$ and the BS$^4$,

• $h_{r,k} \in \mathbb{C}^N$ denotes the conjugate of uplink channel between UT $k$ and the IRS-unit.

Following the fact that the reflecting elements are passive, we can write $\Phi = \text{diag}\{\phi\}$, for some $\phi = [\phi_1, \cdots, \phi_N]^T$ where $\phi_n$ is of the form

$$\phi_n = \beta_n e^{-j\theta_n}$$

(2)

for some $\beta_n \in [0, 1]$ and $\theta_n \in [0, 2\pi)$ denoting the attenuation coefficient and the phase shift applied by the $n$-th reflecting element, respectively. We assume that the attenuation coefficients of all IRS elements are identical. In other words, we consider a case in which $\beta_n$ for $n \in [N]$ are not tunable$^5$. Such an assumption follows from the following two facts:

1) Current technology suggests that implementationally efficient IRSs consist of elements whose attenuation characteristics are not tunable [56]. As the result, practical designs for IRS-assisted communications ignore this degree of freedom$^6$.

2) Due to the long distance between the BS and IRS, the so-called tapering effect is negligible in our setting and can be included in the path-loss model: The tapering effect refers to the non-uniform distribution of attenuation coefficients on the elements of an IRS, due to their different distances and angles of arrival to the transmit array antenna. This effect is tangible in settings in which the IRS is a part of the transmitter architecture, e.g., the scenarios in [24]–[26]. This is not the case in our setting, as we assume that the IRS is a passive element located in a long distance out of the BS site. As the result, the variation of the attenuation coefficients from one IRS element to another is insignificant. This small tapering effect can be ignored at the IRS and be incorporated in the channel path-loss.

Following the above discussions, we set $\beta_n = 1$ for $n \in [N]$.

The received signal $r$ is further given in terms of $x$ as

$$r = T^H x,$$

(3)

$^4$Representing the uplink channel vectors via conjugated vectors is for sake of simplicity.

$^5$This means that the control unit which controls the IRS does not update the attenuation coefficients.

$^6$Although it is theoretically possible to be considered.
where $T^* \in \mathbb{C}^{M \times N}$ contains the coefficients of the uplink channels from the IRS elements to the BS. Consequently, the input-output relation in (1) can be represented as

$$y_k = \tilde{h}_k^H(\phi) x + \vartheta_k,$$

where $\tilde{h}_k(\phi) \in \mathbb{C}^M$ describes the effective end-to-end uplink channel between the $k$-th legitimate UT and the BS, and is given by

$$\tilde{h}_k(\phi) = h_{d,k} + T\Phi h_{r,k}$$

$$= h_{d,k} + \Phi h_{r,k}$$

with $H_k = T\Phi h_{r,k}$. $H_k$ can be regarded as the cascaded effective uplink channel between UT $k$ and the BS through the IRS. We further define the following notation:

$$\tilde{H}(\phi) = [\tilde{h}_1(\phi), \cdots, \tilde{h}_K(\phi)]$$

and refer to it as the legitimate channel matrix.

With a similar approach, the signal received by eavesdropper $j$ is written as

$$z_j = \tilde{g}_j^H(\phi) x + \xi_j,$$

where $\xi_j \sim \mathcal{CN}(0, \mu_j^2)$ is Gaussian noise, and $\tilde{g}_j(\phi) \in \mathbb{C}^M$ describes the effective end-to-end uplink channel from eavesdropper $j$ to the BS being given by

$$\tilde{g}_j(\phi) = g_{d,j} + T\Phi g_{r,j}$$

$$= g_{d,j} + \Phi g_{r,j}$$

In (8c), $g_{d,j}$ and $g_{r,j}$ are the conjugates of the channels corresponding to the path from eavesdropper $j$ to the BS and the IRS-unit, respectively. $G_j$ further represents the effective uplink channel between the $j$-th eavesdropper and the BS through the IRS and is defined as

$$G_j = T\Phi g_{r,j}$$

For sake of brevity, we further define the notations

$$\tilde{G}(\phi) = [\tilde{g}_1(\phi), \cdots, \tilde{g}_J(\phi)],$$

$$\Upsilon_e = \text{diag}\left\{\frac{1}{\mu_1}, \cdots, \frac{1}{\mu_J}\right\}$$

and refer to them as the eavesdropping channel and precision matrix, respectively.
B. Linear Precoding at the BS

The BS constructs the transmit signal from the encoded information symbols of the legitimate UTs, i.e., $s_1, \ldots, s_K$, via linear precoding. Let $w_k$ denote the precoding vector for UT $k$. The transmit signal in this case is given by

$$x = \sum_{k=1}^{K} s_k w_k. \tag{11}$$

We assume that the encoded information symbols are independent and identically distributed (i.i.d.) standard complex Gaussian random variables, i.e., $s \sim \mathcal{CN}(0, I_K)$, where $s = [s_1, \ldots, s_K]$.

In order to restrict the transmit power to the maximum allowed power $P_{\text{max}}$, precoding vectors are scaled, such that

$$\mathbb{E}\{\|x\|^2\} \leq P_{\text{max}}. \tag{12}$$

Defining $W = [w_1, \ldots, w_K]$ as the precoding matrix, the transmit signal is compactly written as $x = Ws$. The power constraint can then be written as

$$\mathbb{E}\{\|Ws\|^2\} \leq P_{\text{max}}. \tag{13}$$

Noting that $\mathbb{E}\{ss^H\} = I_K$, the power constraint reduces to

$$\sum_{k=1}^{K} \|w_k\|^2 \leq P_{\text{max}} \tag{14}$$

or equivalently $\|W\|_F^2 \leq P_{\text{max}}$.

C. Secrecy Performance Metric

The downlink transmission is considered to be secure, if the BS reliably transmits the encoded symbols to the legitimate UTs without allowing any information leakage to the malicious terminals. A rate tuple $(R_1, \ldots, R_K)$ at which secure transmission is guaranteed is called an achievable tuple of secrecy rates, and the convex hull of all achievable tuples is defined as the secrecy capacity region.

In general, the capacity region is not a tractable performance metric for a practical system design. An alternative metric is given by considering a classical inner bound of the secrecy capacity region: A rate tuple $(R_1, \ldots, R_K)$ is achievable with precoding matrix $W$ and vector
of phase-shifts $\phi$, if for $k \in [K]$ we have $R_k \leq R_k^s (W, \phi)$ [57]. The lower bound $R_k^s (W, \phi)$ is given by

$$R_k^s (W, \phi) = [R_k^m (W, \phi) - R_k^e (W, \phi)]^+ \quad (15)$$

where $R_k^m (W, \phi)$ and $R_k^e (W, \phi)$ are defined as follows:

- $R_k^m (W, \phi)$ is a lower bound on the maximum achievable rate to legitimate UT $k$ and is given by

$$R_k^m (W, \phi) = \log (1 + \text{SINR}_k (W, \phi)) \quad (16)$$

where $\text{SINR}_k (W, \phi)$ is the signal-to-interference-plus-noise ratio (SINR) at UT $k$ and is determined as

$$\text{SINR}_k (W, \phi) = \frac{|\hat{h}_k^H (\phi) w_k|^2}{\sum_{i=1,i \neq k} |\hat{h}_i^H (\phi) w_i|^2 + \sigma_k^2} \quad (17)$$

- $R_k^e (W, \phi)$ is an upper bound on the maximum information leakage to the eavesdroppers and is expressed as

$$R_k^e (W, \phi) = \log (1 + \text{ESNR}_k (W, \phi)) \quad (18)$$

with $\text{ESNR}_k (W, \phi)$ being defined as

$$\text{ESNR}_k (W, \phi) = \left\| \Upsilon_e \tilde{G}^H (\phi) w_k \right\|^2 \quad (19)$$

By plugging (16) and (18) into (15), $R_k^s (W, \phi)$ is compactly written as

$$R_k^s (W, \phi) = \left[ \log \left( \frac{1 + \text{SINR}_k (W, \phi)}{1 + \text{ESNR}_k (W, \phi)} \right) \right]^+ \quad (20)$$

The above inner bound is derived by considering the worst-case scenario in which all eavesdroppers cooperate to overhear the secure transmission and are able to cancel out the interference of other legitimate UTs. It is hence clear that larger secrecy rates are also achievable in a practical setting, since such worst-case assumptions are not necessarily fulfilled.

In the sequel, we utilize this inner bound to quantify the secrecy throughput of the system. To this end, we define the weighted secrecy sum-rate as

$$\mathcal{R}_{\text{ssr}} (W, \phi) = \sum_{k=1}^{K} \omega_k R_k^s (W, \phi) \quad (21)$$

for some non-negative real weights $\omega_1, \ldots, \omega_K$ corresponding to the quality-of-services (QoSs) desired for the UTs.
III. JOINTLY OPTIMAL PRECODING AND PHASE-SHIFTING

Considering the problem formulation, the design parameters in this system are the precoding vectors at the BS and phase-shifts applied by the IRS-unit. As a result, the ultimate goal of this work is to design a joint precoding and phase-shifting algorithm that maximizes the weighted secrecy sum-rate.

With respect to the given performance metric, the optimal design for the precoding matrix and phase shifts is given via the following constrained optimization $P_1$:

$$\max_{W,\phi} R_{ssr}^{\text{sec}}(W, \phi) \quad (P_1)$$

subject to

$$\|W\|_F^2 \leq P_{\text{max}} \quad (22a)$$

$$|\phi_n| = 1, \forall n \in [N]. \quad (22b)$$

Here, (22a) ensures that the maximum transmit power is kept below the BS power budget $P_{\text{max}}$, and (22b) restricts the unit modulus constraints imposed by the physical characteristics of the IRS elements.

The optimization problem $P_1$ contains two analytical challenges:

(a) The objective function is not differentiable with respect to the optimization variables. This follows from the fact that the non-negative operator, i.e., $f(x) = [x]^+$, is not differentiable.

(b) Even by replacing the non-negative operators with their arguments in $R_{ssr}^{\text{sec}}(W, \phi)$, $P_1$ is still non-convex. This is due to non-convexity of the objective function and the unit modulus constraints in (22b).

The above challenges indicate that the global optimum of $P_1$ can not be computed via classical convex programming techniques. In fact, it requires a brute-force search that is not a computationally tractable option even for not-so-large dimensions.

We address these issues as follows: First, a variational optimization problem is derived whose objective is differentiable and whose solution coincides with the solution of $P_1$. We then develop a two-tiers algorithm that uses block coordinate descent (BCD) and FP to approximate the solution of the variational problem.

The Variational Problem

Theorem 1 gives an alternative optimization problem with a differentiable objective function whose solution lies on the solution of $P_1$. 

**Theorem 1** (The variational problem). Let $R^*$ be the maximum of the optimization problem $P_1$. Then, we have

$$R^* = \max_{W, \phi, b} \mathcal{R}_{qs} (W, \phi, b) \quad (P_2)$$

subject to $\|W\|^2_F \leq P_{\text{max}}$ \hspace{1cm} (23a)

$|\phi_n| = 1, \ \forall n \in [N],$ \hspace{1cm} (23b)

$b \in [0, 1]^K$ \hspace{1cm} (23c)

where the objective function $\mathcal{R}_{qs} (W, \phi, b)$ is given by

$$\mathcal{R}_{qs} (W, \phi, b) = \sum_{k=1}^{K} \omega_k b_k \log \left( \frac{1 + \text{SINR}_k (W, \phi)}{1 + \text{ESNR}_k (W, \phi)} \right) \quad (24)$$

with $b_k$ denoting the $k$-th entry of $b$.

**Proof.** The equivalency of $P_1$ and $P_2$ is justified by showing that the maximum of $P_1$ bounds the maximum of $P_2$ from below and above. The simultaneous validity of both bounds hence concludes the equivalence of the maxima. The detailed proof is given in Appendix A. $\square$

Theorem 1 addresses the non-differentiability issue of the objective function in $P_1$. Nevertheless, the alternative form in $P_2$ is still non-convex. We address this latter issue in the sequel by developing an iterative algorithm which approximates the solution of $P_2$.

**IV. Analytical Tools From Fractional Programming**

FP is the key analytical tool used in this work. We hence give a quick overview on the main concepts in FP in this section. To this end, we introduce the Lagrangian dual transform for a sum of logarithmic ratios problem and the quadratic transform for a multiple-ratio problem [49], [50]. These are the key results which help us tackling the optimization $P_2$.

**A. Lagrangian Dual and Quadratic Transforms**

The generic form of a *sum of logarithmic ratios* is given by

$$S(x) = \sum_{k=1}^{K} \omega_k \log \left( 1 + \frac{A_k(x)}{B_k(x)} \right). \quad (25)$$

where $\omega_k$ is a non-negative real, $A_k(\cdot) : \mathbb{C}^N \mapsto \mathbb{R}_0^+$ is a non-negative function, and $B_k(\cdot) : \mathbb{C}^N \mapsto \mathbb{R}^+$ is a strictly positive function, for $k \in [K]$. 
Consider the following optimization problem

\[
\max_x S(x)
\]  

subject to \(x \in \mathcal{X}\) \hspace{1cm} (26a)

for some \(\mathcal{X} \subseteq \mathbb{C}^N\) which is non-empty. The Lagrangian dual transform gives an equivalent optimization problem whose both maximum and maximizer recover those given by (26) [50].

**Definition 1** (Lagrangian dual objective). For auxiliary vector \(\mathbf{t} = [t_1, \ldots, t_K] \in \mathbb{R}_0^{+K}\), the Lagrangian dual objective of the sum of logarithmic ratios \(S(x)\) is given by

\[
L(x, \mathbf{t}) = \sum_{k=1}^{K} \omega_k \left( \Xi(t_k) + \frac{(1 + t_k) A_k(x)}{A_k(x) + B_k(x)} \right)
\]  

where the function \(\Xi(\cdot) : \mathbb{R}_0^+ \mapsto \mathbb{R}\) is defined as

\[
\Xi(x) = \log(1 + x) - x.
\]  

In [50, Theorem 3], it is shown that the maximum value of the objective function in (26), as well as the point at which the objective is maximized, are given by the following equivalent optimization problem:

\[
\max_{x, \mathbf{t}} L(x, \mathbf{t})
\]  

subject to \(x \in \mathcal{X}\) and \(\mathbf{t} \in \mathbb{R}_0^{+K}\). \hspace{1cm} (29a)

We refer to this transformed version, as the Lagrangian dual transform of the optimization problem in (26).

Similar to the Lagrangian dual transform, the quadratic transform provides an equivalent optimization problem, when a multiple-ratio function is to be maximized [49]. To illustrate this transform, consider a multiple-ratio function

\[
M(x) = \sum_{k=1}^{K} \frac{|C_k(x)|^2}{D_k(x)},
\]  

in which \(C_k(\cdot) : \mathbb{C}^N \mapsto \mathbb{C}\), and \(D_k(\cdot) : \mathbb{C}^N \mapsto \mathbb{R}^+\) is a strictly positive function, for \(k \in [K]\). We now focus on the following target optimization problem

\[
\max_x M(x)
\]  

subject to \(x \in \mathcal{X}\) \hspace{1cm} (31a)
for some non-empty $\mathcal{X} \subseteq \mathbb{C}^N$. An equivalent optimization problem in this case is given by the quadratic transformed [49].

**Definition 2** (Quadratic equivalent objective). For auxiliary vector $\beta = [\beta_1, \ldots, \beta_K] \in \mathbb{C}^K$, the quadratic equivalent objective of the multiple-ratio function $M(x)$ is given by

$$Q(x, \beta) = \sum_{k=1}^{K} 2 \mathbb{R} \{ \beta_k^* C_k(x) \} - |\beta_k|^2 D_k(x).$$

(32)

Theorem 2 in [49] states that the solution of (31) is given by solving the following equivalent optimization problem:

$$\max_{x, \beta} Q(x, \beta)$$

subject to $x \in \mathcal{X}$ and $\beta \in \mathbb{C}^K$. (33a)

We refer to (33) as the quadratic transform of (31).

V. DEVELOPING AN ITERATIVE ALGORITHM

Starting from $P_2$, we now develop an iterative algorithm which approximates the maximum weighted secrecy sum-rate tractably. The algorithm consists of two tiers: In the first tier, the BCD technique is used to approximate the solution of the joint optimization problem $P_2$ by a cyclic alternation among multiple marginal optimization problems. In the second tier, we use FP and the majorization-maximization (MM) algorithm [58] to address each of these marginal sub-problems.

We start the derivations with the first tier: Considering the optimization problem $P_2$, we group the optimization variables into three blocks; namely, $W$, $\phi$, and $b$. Noting that these blocks are not coupled via the constraints of $P_2$, we use a BCD-type algorithm to approximate the maxima by cyclically alternating among the three marginal problems [59]. In each of these problems, a marginal optimization with respect to one of the blocks is performed while the other two blocks are treated as fixed variables. The alternation among these marginal problems ends when all the blocks converge.

Despite the complexity reduction achieved via the AO technique in the first tier, there exists still a challenge: Two marginal problems deal with a non-convex optimizations. We address this issue in the second tier, where we use the FP and the MM algorithm to approximate the solution of each marginal problem tractably.
A. First Marginal Problem

The first marginal optimization finds the precoding matrix that maximizes the objective $R_{s}\left(W, \phi, b\right)$ while treating $\phi$ and $b$ as fixed variables in $P_2$, i.e., it finds $W^*$ as

$$W^* = \arg\max_{W} R_{s}\left(W, \phi_0, b_0\right) \quad (M_1)$$

subject to $\|W\|^2_F \leq P_{\text{max}}$.

for some fixed $\phi_0$ and $b_0$.

It is straightforwardly seen from (24) that the marginal problem $M_1$ is non-convex. Considering the logarithmic form of the objective function, we use the Lagrangian dual transform to tackle this problem. To this end, we note that the objective function can be decomposed as

$$R_{s}\left(W, \phi_0, b_0\right) = S^m_1(W) + S^e_1(W) \quad (35)$$

where the functions $S^m_1(W)$ and $S^e_1(W)$ are sums of logarithmic ratios and are defined as

$$S^m_1(W) = \sum_{k=1}^{K} \omega_k b_{0k} \log \left(1 + A^m_{1k}(W) \right)$$

$$S^e_1(W) = \sum_{k=1}^{K} \omega_k b_{0k} \log \left( \frac{1}{B^e_{1k}(W)} \right). \quad (36a, 36b)$$

with $A^m_{1k}(W)$, $B^m_{1k}(W)$ and $B^e_{1k}(W)$ being

$$A^m_{1k}(W) = |\tilde{h}_k^H(\phi_0)w_k|^2, \quad (37a)$$

$$B^m_{1k}(W) = \sum_{i=1, i\neq k}^{K} |\tilde{h}_i^H(\phi_0)w_i|^2 + \sigma_k^2, \quad (37b)$$

$$B^e_{1k}(W) = 1 + \|\Upsilon e \tilde{G}^H(\phi_0)w_k\|^2_F. \quad (37c)$$

Although $S^m_1(W)$ is of the standard form given in (25), the function $S^e_1(W)$ does not fulfill the non-negativity constraint of the nominator: One can rewrite $S^e_1(W)$ as

$$S^e_1(W) = \sum_{k=1}^{K} \omega_k b_{0k} \log \left(1 + \frac{1}{B^e_{1k}(W)} \right). \quad (38)$$

However, in this case $1 - B^e_{1k}(W) \leq 0$ which violates the necessary condition of having a non-negative nominator. To overcome this issue, we note that

$$B^e_{1k}(W) \leq 1 + \|\Upsilon e \tilde{G}^H(\phi_0)\|^2_F \|w_k\|^2. \quad (39)$$
Since the power constraint enforces $\|w_k\|^2 \leq P_{\text{max}}$, we could conclude that $B^e_{1k}(W) \leq B^e_{1U}$ with\footnote{Note that the upper bound is fixed in terms of $W$.}

$$B^e_{1U} = 1 + \|Y_0 \hat{G}^n( \phi_0) \|^2_P P_{\text{max}}.$$  

Using this upper bound, we can rewrite $S^m_1(W)$ as

$$S^e_1 (W) = \sum_{k=1}^{K} \omega_k b_{0k} \log \left( \frac{1}{B^e_{1k}(W)} \right)$$

$$= \sum_{k=1}^{K} \omega_k b_{0k} \log \left( \frac{B^e_{1U}}{B^e_{1k}(W)} \right) - \log B^e_{1U}$$

$$= \sum_{k=1}^{K} \omega_k b_{0k} \log \left( 1 + \frac{A^e_{1k}(W)}{B^e_{1k}(W)} \right) - \log B^e_{1U}$$

where we define

$$A^e_{1k}(W) = B^e_{1U} - B^e_{1k}(W).$$

In this alternative representation, $A^e_{1k}(W) \geq 0$ which satisfies the non-negativity constraint required for using the Lagrangian dual transform.

By defining $\hat{S}^e_1(W)$ as

$$\hat{S}^e_1 (W) = \sum_{k=1}^{K} \omega_k b_{0k} \log \left( 1 + \frac{A^e_{1k}(W)}{B^e_{1k}(W)} \right),$$

we could conclude that

$$W^* = \arg\max_W S^m_1(W) + \hat{S}^e_1(W)$$

$$\text{(M}_1\text{)}$$

subject to $\|W\|^2_F \leq P_{\text{max}}.$

The optimization problem $\hat{M}_1$ is of the standard form given in (26). Using the Lagrangian dual transform, we have

$$\left(W^*, t^*, \alpha^*\right) = \arg\max_{W, t, \alpha} L^n_1(W, t) + L^e_1(W, \alpha)$$

$$\text{(L}_1\text{)}$$

subject to $\|W\|^2_F \leq P_{\text{max}}$  

$$t, \alpha \in \mathbb{R}_0^+K$$

\footnote{Note that the upper bound is fixed in terms of $W$.}
where $L_m^1(W, t)$ and $L_e^1(W, \alpha)$ are the Lagrangian dual objectives of $S_m^1(W)$ and $\bar{S}_1^e(W)$, respectively, and are given by

$$L_m^1(W, t) = \sum_{k=1}^{K} \omega_k b_{0k} \left( \Xi (t_k) + \frac{(1 + t_k) A_{mk}^1(W)}{A_{mk}^1(W) + B_{bk}^m(W)} \right)$$ (46a)

$$L_e^1(W, \alpha) = \sum_{k=1}^{K} \omega_k b_{0k} \left( \Xi (\alpha_k) + \frac{(1 + \alpha_k) A_{ek}^1(W)}{B_{ek}^m(W)} \right)$$ (46b)

with $t_k$ and $\alpha_k$ being the $k$-th entry of $t$ and $\alpha$, respectively.

The Lagrangian dual transform of the marginal problem is a joint optimization with a non-convex objective function. To tackle this issue, we note that the constraints on block variable $W$ and $(t, \alpha)$ decouple in $L_1$. We hence utilize BCD once again and solve the Lagrangian dual transform by cyclically alternating between the following sub-problems:

1) Update $W$ for fixed $t = \bar{t}$ and $\alpha = \bar{\alpha}$ as

$$W = \arg\max_W L_m^1(W, \bar{t}) + L_e^1(W, \bar{\alpha})$$ \hfill ($L_1^A$)

subject to $\|W\|_F^2 \leq P_{\text{max}}$. (47a)

2) Update $t$ and $\alpha$ for a fixed $W = \bar{W}$ as

$$(\bar{t}, \bar{\alpha}) = \arg\max_{t, \alpha} L_m^1(\bar{W}, t) + L_e^1(\bar{W}, \alpha)$$ \hfill ($L_1^B$)

subject to $t, \alpha \in \mathbb{R}_0^{+K}$. (48a)

The sub-problem $L_1^A$ optimized the objective with respect to $W$. By dropping those terms which are constant in $W$, we can rewrite $L_1^A$ as

$$\bar{W} = \arg\max_W \sum_{k=1}^{K} \frac{|C_{1k}^m(W)|^2}{D_{1k}^m(W)} + \|C_{1k}^e(W)\|^2$$ \hfill ($L_1^A$)

subject to $\|W\|_F^2 \leq P_{\text{max}}$. (49a)

where $C_{1k}^m(W)$, $D_{1k}^m(W)$, and $C_{1k}^e(W)$ are defined as

$$C_{1k}^m(W) = \sqrt{\omega_k b_{0k} (1 + t_k)} \Gamma_k (\phi_0) w_k$$, \hfill (50a)

$$D_{1k}^m(W) = A_{1k}^m(W) + B_{1k}^m(W)$$, \hfill (50b)

$$C_{1k}^e(W) = \sqrt{\frac{\omega_k b_{0k} (1 + \alpha_k)}{B_{1U}^e}} \sqrt{A_{1k}^e(W)}$$.

(50c)
The optimization problem $\hat{L}_A$ is a multiple-ratio maximization whose quadratic transform is given by

$$\max_{W, \beta, \gamma} Q_m^m(W, \beta) + Q_e^e(W, \gamma)$$

subject to $\|W\|_F^2 \leq P_{\text{max}}$

$$\beta, \gamma \in \mathbb{C}^K$$

with quadratic equivalent objectives

$$Q_m^m(W, \beta) = \sum_{k=1}^K 2 \Re \{ \beta_k^* C_{1k}^m(W) \} - |\beta_k|^2 D_{1k}^m(W)$$

$$Q_e^e(W, \gamma) = \sum_{k=1}^K 2 \Re \{ \gamma_k^* C_{1k}^e(W) \} - |\gamma_k|^2.$$

Here, $\beta_k$ and $\gamma_k$ denote the $k$-th entry of $\beta$ and $\gamma$, respectively.

The quadratic transform $Q_1$ can be directly solved. Alternatively, one can use the BCD approach once again. In this case, we consider two blocks of variables: $(W, \gamma)$ and $\beta$. The solution is then found by alternating between the following two sub-problems:

(A) Find optimal $\beta$ for $W = \bar{W}$ and $\gamma = \bar{\gamma}$ as

$$\bar{\beta} = \arg\max_{\beta} Q_m^m(\bar{W}, \beta) + Q_e^e(\bar{W}, \bar{\gamma})$$

subject to $\beta \in \mathbb{C}^K$.

$Q_A^A$ is a standard quadratic optimization with solution

$$\bar{\beta}_k = \frac{C_{1k}^m(\bar{W})}{D_{1k}^m(\bar{W})}.$$  

(B) Find optimal $W$ and $\gamma$ for $\beta = \bar{\beta}$ as

$$(\bar{W}, \bar{\gamma}) = \arg\max_{W, \gamma} Q_m^m(W, \bar{\beta}) + Q_e^e(W, \gamma)$$

subject to $\|W\|_F^2 \leq P_{\text{max}}$ and $\gamma \in \mathbb{C}^K$.

Similar to $Q_A^A$, $Q_B^A$ is a standard quadratic optimization whose solution is given by

$$\bar{W} = [\bar{w}_1, \ldots, \bar{w}_K],$$

$$\bar{\gamma} = [C_{11}^e(\bar{W}), \ldots, C_{1K}^m(\bar{W})].$$
In (56a), \( \vec{w}_k \) for \( k \in [K] \) is
\[
\vec{w}_k = \rho_k \Gamma_k^{-1} \tilde{h}_k(\phi_0)
\]  

(57)

where \( \Gamma_k \) and \( \rho_k \) are given by
\[
\Gamma_k = \tilde{H}(\phi_0) \Upsilon_m^2 \tilde{H}^H(\phi_0) + \tau_k \tilde{G}(\phi_0) \Upsilon_e^2 \tilde{G}^H(\phi_0) + \lambda_P I_M,
\]

(58a)

\[
\rho_k = \sqrt{\omega_k b_{0k} (1 + \bar{\alpha}_k) \beta_k}
\]

(58b)

with \( \Upsilon_m \) and \( \tau_k \) being defined as
\[
\Upsilon_m = \text{diag}\{ |\beta_1|, \ldots, |\beta_k| \},
\]

(59a)

\[
\tau_k = \frac{\omega_k b_{0k} (1 + \bar{\alpha}_k)}{B_{eU}}.
\]

(59b)

and some \( \lambda_P \) being calculated according to the power constraint\(^8\).

Noting that \( \vec{W} \) is fixed in terms of \( \vec{\gamma} \), the auxiliary vector \( \vec{\gamma} \) can be ignored in the BCD algorithm. Hence, the solution to \( L_1^A \) is given by alternating between (56a) and (54).

**Remark 2.** The precoding matrix \( \vec{W} \) inverts the main channels while aligning the nulls of the beams with the eavesdropping channels. This is the so-called secure regularized zero-forcing (SRZF) precoding technique which has been proposed and studied in [63]. Unlike the generic form of SRZF precoding, \( \vec{W} \) does not need to be tuned. This is due to the fact that \( \vec{W} \) is directly derived from secrecy sum-rate maximization.

We now discuss the solution to the sub-problem \( L_1^B \). Following the concavity of the objective function and its decoupled form, it is straightforward to show that
\[
\bar{t}_k = \frac{A_{1k}^m (\vec{W})}{B_{1k}^m (\vec{W})},
\]

(60a)

\[
\bar{\alpha}_k = \frac{A_{1k}^e (\vec{W})}{B_{1k}^e (\vec{W})}.
\]

(60b)

Consequently, the solution to the dual problem \( L_1 \) is given by alternating between the solution of the sub-problem \( L_1^A \) and the updates in (60). Once the alternations converge, the solution to the first marginal problem is given by \( \vec{W} \).

\(^8\)\( \lambda_P \) can be tuned optimally via various algorithms, e.g., bi-section search [16], [60]. Alternatively, one can solve \( Q_1^B \) directly via a convex programming solver, e.g., the CVX package in MATLAB [61], [62].
B. Second Marginal Problem

In the second marginal optimization, we determine a vector of phase-shifts which maximizes the objective \( R_{ssr}^q (W, \phi, b) \) assuming that \( W \) and \( b \) are fixed. Hence, the corresponding optimization is

\[
\phi^* = \arg\max_\phi R_{ssr}^q (W_0, \phi, b_0) \tag{M_2}
\]

subject to \(|\phi_n| = 1, \forall n \in [N] \),

for some fixed \( W_0 \) and \( b_0 \).

In its initial form, the marginal problem \( M_2 \) is intractable due to the following issues:

1) The objective function is non-convex.
2) The constraint is unit-modulus.

Considering the fractional form of the objective function, the first issue is addressed by transforming the marginal problem \( M_2 \) to a quadratic optimization using the Lagrangian dual and the quadratic transforms. Given the transformed optimization, we further overcome the second issue via the MM method. The detailed derivations are presented in the sequel.

We start the derivations by rewriting the objective function as a sum of logarithmic ratios. Similar to the marginal problem \( M_1 \), we can write

\[
R_{ssr}^q (W_0, \phi, b_0) = S_2^m (\phi) + S_2^e (\phi) \tag{62}
\]

where the functions \( S_2^m (\phi) \) and \( S_2^e (\phi) \) are defined as

\[
S_2^m (\phi) = \sum_{k=1}^{K} \omega_k b_{0k} \log \left( 1 + \frac{A_{2k}^m (\phi)}{B_{2k}^m (\phi)} \right), \tag{63a}
\]

\[
S_2^e (\phi) = \sum_{k=1}^{K} \omega_k b_{0k} \log \left( \frac{1}{B_{2k}^e (\phi)} \right). \tag{63b}
\]

Here, \( A_{2k}^m (\phi), B_{2k}^m (\phi), \) and \( B_{2k}^e (\phi) \) are given by

\[
A_{2k}^m (\phi) = |\tilde{h}_k^H (\phi) w_{0k}|^2, \tag{64a}
\]

\[
B_{2k}^m (\phi) = \sum_{i=1, i \neq k}^{K} |\tilde{h}_k^H (\phi) w_{0i}|^2 + \sigma_k^2, \tag{64b}
\]

\[
B_{2k}^e (\phi) = 1 + \text{ESNR}_k (W_0, \phi) \tag{64c}
\]

where \( w_{0k} \) denotes the \( k \)-th column of \( W_0 \).
In order to use the Lagrangian dual transform, we need to convert $S^e_2 (\phi)$ to the standard form given in (25). To this end, we substitute $g_j (\phi) = g_{d,j} + G_j \phi$ into (19) and write

$$
\text{ESNR}_k (W, \phi) = \sum_{j=1}^J \frac{1}{\mu^2_j} |g_{d,j}^H w_k + \phi^H G_j^H w_k|^2
$$

(65a)

\[ \leq \sum_{j=1}^J \frac{1}{\mu^2_j} (|g_{d,j}^H w_k| + |\phi^H G_j^H w_k|)^2 \]  

(65b)

\[ \leq \sum_{j=1}^J \frac{1}{\mu^2_j} (|g_{d,j}^H w_k| + \| \phi^H G_j^H w_k \|^2)^2 \]  

(65c)

\[ = \sum_{j=1}^J \frac{1}{\mu^2_j} (|g_{d,j}^H w_k| + \sqrt{N} \| G_j^H w_k \|)^2 \]  

(65d)

where $\star$ comes from the triangle inequality, and $\dagger$ follows from the fact that $|\phi_n| = 1$. Defining the upper-bound $B_{2Uk}^e$ for $B_{2k}^e (\phi)$ as

$$
B_{2Uk}^e = 1 + \sum_{j=1}^J \frac{1}{\mu^2_j} \left( |g_{d,j}^H w_{0k}| + \sqrt{N} \| G_j^H w_{0k} \| \right)^2 ,
$$

(66)

we can rewrite $S^e_2 (\phi)$ as

$$
S^e_2 (\phi) = \hat{S}^e_2 (\phi) - \sum_{k=1}^K \omega_k b_k \log B_{2Uk}^e
$$

(67)

where $\hat{S}^e_2 (\phi)$ is given by

$$
\hat{S}^e_2 (\phi) = \sum_{k=1}^K \omega_k b_k \log \left( 1 + \frac{A_{2k}^e (\phi)}{B_{2k}^e (\phi)} \right)
$$

(68)

with $A_{2k}^e (\phi) = B_{2Uk}^e - B_{2k}^e (\phi) \geq 0$.

Given this alternative representation of the objective, the marginal problem $\mathcal{M}_2$ is rewritten as

$$
\phi^* = \arg\max_\phi S^m_2 (\phi) + \hat{S}^e_2 (\phi) \quad (\hat{\mathcal{M}}_2) \tag{69a}
$$

subject to $|\phi_n| = 1, \forall n \in [N]$  

which is of the standard form in (26). Using the Lagrangian dual transform, we could find $\phi^*$ from the transformed version

$$
(\phi^*, \mathbf{q}^*, \psi^*) = \arg\max_{\phi, \mathbf{q}, \psi} L^m_2 (\phi, \mathbf{q}) + L^e_2 (\phi, \psi) \quad (\mathcal{L}_2) \tag{70a}
$$

subject to $|\phi_n| = 1, \forall n \in [N]$  

$$
\mathbf{q}, \psi \in \mathbb{R}_0^{+K} \tag{70b}
$$
where the Lagrangian dual objectives are

\[
L_m^2 (\phi, q) = \sum_{k=1}^{K} \omega_k b_{0k} \left( \Xi (q_k) + (1 + q_k) \frac{A_{2k}^m (\phi)}{A_{2k}^m (\phi) + B_{2k}^m (\phi)} \right),
\]

\[
L_e^2 (\phi, \psi) = \sum_{k=1}^{K} \omega_k b_{0k} \left( \Xi (\psi_k) + (1 + \psi_k) \frac{A_{2k}^e (\phi)}{B_{2k}^e (\phi)} \right),
\]

with \(q_k\) and \(\psi_k\) denoting entry \(k\) of \(q\) and \(\psi\), respectively.

To address the Lagrangian dual problem, we follow the same BCD-based approach as for the first marginal problem: Starting from an initial point, we alternate between the following two sub-problems:

1) Update \(\phi\) for fixed \(q = \bar{q}\) and \(\psi = \bar{\psi}\) as

\[
\bar{\phi} = \arg\max_{\phi} L_m^2 (\phi, \bar{q}) + L_e^2 (\phi, \bar{\psi})
\]  \( (L_2^A) \)

subject to \(|\phi_n| = 1, \forall n \in [N] \).

2) Update \(q\) and \(\psi\) for fixed \(\phi = \bar{\phi}\) as

\[
(\bar{q}, \bar{\psi}) = \arg\max_{q, \psi} L_m^2 (\bar{\phi}, q) + L_e^2 (\bar{\phi}, \psi)
\]  \( (L_2^B) \)

subject to \(q, \psi \in \mathbb{R}_0^{+K} \).

Noting that the sub-problem \(L_2^A\) is to maximize a multiple-ratio term, we use the quadratic transform to find the equivalent quadratic problem. After dropping the constant terms and some standard lines of derivation the equivalent quadratic problem becomes

\[
\max_{\phi, f, \varpi} Q_m^m (\phi, f) + Q_e^e (\phi, \varpi)
\]  \( (Q_2) \)

subject to \(|\phi_n| = 1, \forall n \in [N] \)

\( f, \varpi \in \mathbb{C}^K \)

for the quadratic equivalent objectives

\[
Q_m^m (\phi, f) = \sum_{k=1}^{K} 2 \Re \{ f_k^* C_{2k}^m (\phi) \} - |f_k|^2 D_{2k}^m (\phi),
\]

\[
Q_e^e (\phi, \varpi) = \sum_{k=1}^{K} 2 \Re \{ \varpi_k^* C_{2k}^e (\phi) \} - |\varpi_k|^2,
\]
where \( f_k \) and \( \varpi_k \) are entry \( k \) of \( f \) and \( \varpi \), respectively, and \( C_{2k}^m(\phi) \), \( D_{2k}^m(\phi) \), and \( C_{2k}^e(\phi) \) are defined as

\[
C_{2k}^m(\phi) = \sqrt{\omega_k b_{0k}(1+\bar{q}_k)} \left( h_{d,k}^H w_{0k} + \phi^H H_k^H w_{0k} \right),
\]

\[
D_{2k}^m(\phi) = \| h_{d,k}^H W_0 + \phi^H H_k^H W_0 \|^2 + \sigma_k^2,
\]

\[
C_{2k}^e(\phi) = \sqrt{\omega_k b_{0k} \left( 1 + \bar{\psi}_k \right)} B_{2U_k}^e(\phi).
\]

As the constraints in \( Q_2 \) decouple, we find the solution via alternation between the following two steps

(A) Find optimal \( f \) for \( \phi = \bar{\phi} \) and \( \varpi = \bar{\varpi} \) as

\[
\bar{\phi} = \arg\max_f Q_{2k}^m(\bar{\phi}, f) + Q_{2}^e(\bar{\phi}, \bar{\varpi})
\]

subject to \( f \in \mathbb{C}^K \).

The solution to this marginal optimization is

\[
\bar{f}_k = \frac{C_{2k}^m(\bar{\phi})}{D_{2k}^m(\bar{\phi})}.
\]

(B) Find optimal \( \phi \) and \( \varpi \) for \( f = \bar{f} \) as

\[
(\phi^*, \varpi) = \arg\max_{\phi, \varpi} Q_{2k}^m(\phi, \bar{f}) + Q_{2}^e(\phi, \varpi)
\]

subject to \( |\phi_n| = 1, \; \forall n \in [N] \)

\[
\varpi \in \mathbb{C}^K.
\]

It is straightforward to show that the variable \( \varpi \) in \( Q_2^B \) does not participate in the update of step (A), and hence \( Q_2^B \) can be replaced with

\[
\phi^* = \arg\max_{\phi} Q_{2k}^m(\phi, \bar{f}) + \frac{\omega_k b_{0k}(1+\bar{\psi}_k)}{B_{2U_k}^e} A_{2k}^e(\phi)
\]

subject to \( |\phi_n| = 1, \; \forall n \in [N] \).

Despite the quadratic objective of \( Q_2^B \), the optimization is still intractable due to the unit-modulus constraint. There are various approaches by which \( \phi^* \) can be approximated tractably; see for instance [58], [64], and [65]. In the following, we consider the algorithm \( Q_{MM}(\cdot) \) which uses MM method to determine the approximation

\[
\tilde{\phi} = Q_{MM}(\bar{f}, \bar{q}, \bar{\psi}, W_0)
\]
iteratively. The algorithm is illustrated in Algorithm 1. In this algorithm, \( U \) and \( v \) are determined in terms of \( \tilde{f} \) as

\[
U = \sum_{k=1}^{K} |\tilde{f}_k|^2 \mathbf{H}_k^H \mathbf{W}_0 \mathbf{W}_0^H \mathbf{H}_k + \sum_{k=1}^{K} \sum_{j=1}^{J} \frac{\kappa_k}{\mu_j^2} G_j^H \mathbf{w}_{0k} \mathbf{w}_{0k}^H G_j, \quad (81a)
\]

\[
v = \sum_{k=1}^{K} |\tilde{f}_k|^2 \mathbf{H}_k^H \mathbf{W}_0 \mathbf{W}_0^H \mathbf{h}_{d,k} - \eta_k \mathbf{H}_k^H \mathbf{w}_{0k} + \sum_{k=1}^{K} \sum_{j=1}^{J} \frac{\kappa_k}{\mu_j^2} G_j^H \mathbf{w}_{0k} \mathbf{w}_{0k}^H g_{d,j}, \quad (81b)
\]

where \( \kappa_k \) and \( \eta_k \) are given by

\[
\kappa_k = \frac{\omega_k b_{0k} (1 + \bar{\psi}_k)}{B_{2U/k}}, \quad (82a)
\]

\[
\eta_k = \tilde{f}_k^* \sqrt{\omega_k b_{0k} (1 + \bar{\psi}_k)}. \quad (82b)
\]

Furthermore, \( u_n^H \) and \( v_n \) denote the \( n \)-th row of \( U \) and \( v \), respectively, and \( \lambda_{\text{max}} \) represents the maximum eigenvalue of \( U \). The detailed derivations are given in Appendix B. Following the classical approach, e.g., the approach in [66], the convergence of Algorithm 1 can be shown. This means that Algorithm 1 in each iteration updates the phase-shift, such that the objective in \( \hat{Q}_2^B \) evolves in a non-decreasing fashion\(^9\). Due to the similarity, we drop the convergence analysis, and refer the interested reader to [66] and the references therein.

**Algorithm 1** The MM-based algorithm \( Q_{\text{MM}} (\cdot) \)

**Input:** \( \tilde{f}, \bar{q}, \bar{\psi} \) and \( \mathbf{W}_0 \)

**Initialization:** Set a feasible initial point \( \phi \), and define

\[
\mathcal{R}_{\text{MM}}^\text{ssr} (\phi) = Q_{2}^\text{imi} (\phi, \tilde{f}) + \frac{\omega_k b_{0k} (1 + \bar{\psi}_k)}{B_{2U/k}} A_{2k}^\text{e} (\phi).
\]

Calculate \( U \) and \( v \) from (81), and let \( \lambda_{\text{max}} \) be the maximum eigenvalue of \( U \).

1: if \( \mathcal{R}_{\text{MM}}^\text{ssr} (\phi) \) has not converged then
2: Set \( \hat{\phi}_n = u_n^H \phi + v_n - \lambda_{\text{max}} \phi_n \) for \( n \in [N] \)
3: Update \( \phi_n = -\frac{\hat{\phi}_n}{|\phi_n|} \) for \( n \in [N] \), and go back to line 1
4: end if

**Remark 3.** It is worth mentioning that \( Q_{\text{MM}} (\cdot) \) does not necessarily results in the best approximation possible to be determined tractably. One can hence also consider other possible approaches

---

\(^9\)More precisely, the objective at the end of each iteration is larger or equal to the objective value in the previous iteration. Noting that the objective has a bounded maximum, this concludes that the algorithm converges to a fixed-point.
for approximating $\phi^*$. An example of an alternative algorithm based on the BCD method is given in Appendix C.

By alternating between (80) and (77b), the solution to the sub-problem $L_2^A$ is determined. It is further straightforward to show that the solution to $L_2^B$, when $\phi = \bar{\phi}$, is given by

$$
\bar{q}_k = \frac{A_{2k}^m (\bar{\phi})}{B_{2k}^m (\bar{\phi})}, \quad (83a)
$$

$$
\bar{\psi}_k = \frac{A_{2k}^e (\bar{\phi})}{B_{2k}^e (\bar{\phi})}. \quad (83b)
$$

C. Third Marginal Problem

The third marginal optimization determines the auxiliary vector $b$ which maximizes $R_{q_1}^{\text{srr}} (W, \phi, b)$ for fixed $W$ and $b$. The corresponding optimization is thus given by

$$
b^* = \arg \max_b R_{q_1}^{\text{srr}} (W_0, \phi_0, b) \quad (M_3)
$$

subject to $b \in [0, 1]^K$ \hfill (84a)

for some fixed $W_0$ and $\phi_0$.

$M_3$ is a linear program whose solution is given by

$$
\bar{b}_k = \begin{cases} 
1 & \text{SINR}_k (W_0, \phi_0) > \text{ESNR}_k (W_0, \phi_0) \\
0 & \text{Otherwise}
\end{cases} \quad (85)
$$

This means that in each iteration, the objective includes only those UTs whose achievable secrecy rates are non-zero.

VI. Iterative Algorithms with Reduced Complexity

The BCD method developed in Section V approximates the jointly optimal design using a multi-tier BCD-based algorithm: The algorithm alternates among three marginal problems. The solution to each marginal problem is further determined via the BCD method in which some inner loops alternate among multiple sub-problems. Some of these sub-problems are further solved via alternation among multiple other sub-problems. For instance, the solution to the first marginal problem is given by alternating between the two sub-problems $L_1^A$ and $L_1^B$, and the solution of $L_1^A$, is given by alternating between the update rules in (56a) and (54). Consequently, in each iteration of the first inner loop$^{10}$, the second inner loop$^{11}$ should run for several iterations,

$^{10}$By the first inner loop, we mean alternating between $L_1^A$ and $L_1^B$.

$^{11}$By the second inner loop, we mean alternating between (56a) and (54).
such that the inner BCD-based algorithm converges.

The multi-tier nature of the proposed algorithm yields high computational complexity. In this section, we address this issue by proposing two iterative algorithms with reduced complexity. These algorithms reduce the computational complexity by merging multiple inner loops of different BCD tiers. We show that these algorithms converge to their fixed-points in a non-decreasing fashion, meaning that the updated objective in each iteration is necessarily greater than or equal to the objective calculated in the previous iteration.

A. Two-Tiers Algorithm

The first iterative algorithm comprises of two tiers. The inner tier corresponds to the marginal problems and contains the merged version of inner loops. The algorithm uses the decoupling of sub-problems and perform the updates in parallel. The inner loop of iterations in the second marginal problem, by which the phase-shifts are updated, is however not merged. This leads to a moderate computational complexity.

1) Statement of the Algorithm: The algorithm is represented in Algorithm 2. This algorithm follows the standard BCD approach by alternating among the marginal problems. Here, $\mathcal{A}_1 (\cdot)$ and $\mathcal{A}_2 (\cdot)$ refer to inner loops which approximate the solution of first and second marginal problems, respectively. These loops are presented in Algorithms 3 and 4.

**Algorithm 2** Two-Tiers Algorithm for Precoding and Phase-Shift Tuning

**Initialization:** Set $\mathbf{W}$ and $\phi$ to some initial values.

1. **if** $\mathcal{R}_{\text{ssr}} (\mathbf{W}, \phi)$ has not converged **then**
2. Update $\mathbf{W}$ as $\mathbf{W} = \mathcal{A}_1 (\phi, b)$
3. Update $\phi$ as $\phi = \mathcal{A}_2 (\mathbf{W}, b)$
4. Update $b_k = 1 \{\text{SINR}_k (\mathbf{W}, \phi) > \text{ESNR}_k (\mathbf{W}, \phi)\}$ for $k \in [K]$
5. Go back to line 1
6. **end if**

Algorithm 3 illustrates the inner loop for the first marginal problem. In this algorithm, the inner loops of the sub-problems are merged following the intuitive decoupling of these problems. The convergence of this algorithm is discussed in Lemma 1 in Section VI-A2.

The inner loop $\mathcal{A}_2 (\cdot)$ is presented in Algorithm 4. Here, all the BCD loops are merged into a single loop. The MM loop, by which the phase-shifts are updated is however not merged. This
Algorithm 3 Inner Loop for the First Marginal Problem $A_1(\cdot)$

**Input:** $\phi$ and $b$

**Initialization:** Set $\bar{W}$ to some initial value and define marginal $R^srf_{\text{M}}(W) = R^srf_{\text{q}}(W, \phi, b)$.

1. if $R^srf_{\text{M}}(W)$ has not converged then
   2. Update $\tilde{\beta}_k$ via (54) for $k \in [K]$
   3. Update $\tilde{t}_k$ via (60a) for $k \in [K]$
   4. Update $\tilde{\alpha}_k$ via (60b) for $k \in [K]$
   5. Update $\tilde{w}_k$ via (57) for $k \in [K]$
   6. Set $W_{\text{new}} = [\tilde{w}_1, \ldots, \tilde{w}_K]$
   7. Update $\bar{W} = W_{\text{new}}$ and go back to line 1
   8. end if

means that $A_2(\cdot)$ contains another inner loop, i.e., $Q_{\text{MM}}(\cdot)$. As the result, Algorithm 2 yields moderate computational complexity $Q_{\text{MM}}(\cdot)$ requires larger number of iterations to converge. The convergence of Algorithm 4 is discussed in Lemma 2 in Section VI-A2.

Algorithm 4 Inner Loop for the Second Marginal Problem $A_2(\cdot)$

**Input:** $W$ and $b$

**Initialization:** Set $\bar{\phi}$ to some initial value and define marginal $R^srf_{\text{M}}(\phi) = R^srf_{\text{q}}(W, \phi, b)$.

1. if $R^srf_{\text{M}}(\phi)$ has not converged then
   2. Update $\bar{f}_k$ via (77b) for $k \in [K]$
   3. Update $\bar{q}_k$ via (83a) for $k \in [K]$
   4. Update $\bar{\psi}_k$ via (83b) for $k \in [K]$
   5. Set $\phi_{\text{new}} = Q_{\text{MM}}(\bar{f}, \bar{q}, \bar{\psi}, W)$
   6. Update $\bar{\phi} = \phi_{\text{new}}$ and go back to line 1
   7. end if

2) *Convergence Analysis:* We now discuss the convergence of Algorithm 2. We show that using this two-tiers algorithm, the original objective, i.e., $R^srf(W, \phi)$, evolve non-decreasingly. This means that the value of the objective function after each iteration would be greater or equal to its value in the previous iteration. Since the objective has a bounded maximum, this result
indicates that this algorithm converges to a fixed-point after a certain number of iterations\textsuperscript{12}.

We start our analysis by showing the convergence of the inner loops $A_1 (\cdot)$ and $A_2 (\cdot)$ in the following lemmas:

**Lemma 1.** *For the inner loop $A_1 (\bar{\phi}, \bar{b})$, we have*

\[
R_{qs}^\text{ssr} \left( \bar{W}, \bar{\phi}, \bar{b} \right) \leq R_{qs}^\text{ssr} \left( W_{\text{new}}, \bar{\phi}, \bar{b} \right). \tag{86}
\]

*Proof.* The proof is given in Appendix D.

Lemma 1 indicates that in each iteration, the updated precoding matrix returned a larger value of the transformed objective function $R_{qs}^\text{ssr} (\cdot)$. As the result, one can conclude that after a certain number of iterations, $A_1 (\cdot)$ must converge. This result is further stated for $A_2 (\cdot)$ in Lemma 2.

**Lemma 2.** *Assume that in iteration $t \geq 1$ of the inner loop $A_2 (\bar{W}, \bar{b})$, $Q_{MM} (\cdot)$ iterates for $T_Q (t)$ iterations. For any positive integer sequence of $T_Q (t)$, we have*

\[
R_{qs}^\text{ssr} \left( \bar{W}, \bar{\phi}, \bar{b} \right) \leq R_{qs}^\text{ssr} \left( W, \phi_{\text{new}}, \bar{b} \right). \tag{87}
\]

*This inequality further holds for any alternative of $Q_{MM} (\cdot)$ which converges to its fixed-point in a non-decreasing fashion.*

*Proof.* The proof is given in Appendix E.

Using Lemmas 1 and 2, we state the following result on the convergence of Algorithm 2:

**Theorem 2.** *Let $W^{(t)}$ and $\phi^{(t)}$ be the precoding matrix and vector of phase-shifts updated at the end of iteration $t$ in Algorithm 2. Assume that $A_1 (\cdot)$ and $A_2 (\cdot)$ iterate for $T_1 (t)$ and $T_2 (t)$, respectively, in iteration $t$ of Algorithm 2. For any positive integer sequence of $T_1 (t)$ and $T_2 (t)$, we have*

\[
R_{qs}^\text{ssr} \left( W^{(t)}, \phi^{(t)} \right) \leq R_{qs}^\text{ssr} \left( W^{(t+1)}, \phi^{(t+1)} \right).
\]

*The inequality holds for any alternative of $Q_{MM} (\cdot)$ in the inner loop $A_2 (\cdot)$ which converges to its fixed-point in a non-decreasing fashion and iterates in each iteration of $A_2 (\cdot)$ for $T \geq 1$ iterations.*

\textsuperscript{12}This fixed-point is however not necessarily the global or a local maximum and is only an approximation.
Proof. Let \( b^{(t)} \) denote the updated auxiliary variable in iteration \( t \). Considering Algorithm 2, we can write

\[
b^{(t)}_k = \begin{cases} 
1 & \text{SINR}_k \left( W^{(t)}, \phi^{(t)} \right) > \text{ESNR}_k \left( W^{(t)}, \phi^{(t)} \right) \\
0 & \text{Otherwise}
\end{cases}
\]  

(88)

This concludes that

\[
\mathcal{R}_{q}^{\text{ssr}} \left( W^{(t)}, \phi^{(t)}, b^{(t)} \right) = \sum_{k=1}^{K} \omega_k b^{(t)}_k \log \left( \frac{1 + \text{SINR}_k \left( W^{(t)}, \phi^{(t)} \right)}{1 + \text{ESNR}_k \left( W^{(t)}, \phi^{(t)} \right)} \right)
\]

(89a)

\[
= \sum_{k=1}^{K} \omega_k \left[ \log \left( \frac{1 + \text{SINR}_k \left( W^{(t)}, \phi^{(t)} \right)}{1 + \text{ESNR}_k \left( W^{(t)}, \phi^{(t)} \right)} \right) \right]^{+}
\]

(89b)

\[
= \mathcal{R}_{q}^{\text{ssr}} \left( W^{(t)}, \phi^{(t)} \right)
\]

(89c)

As the result, we can write

\[
\mathcal{R}_{q}^{\text{ssr}} \left( W^{(t)}, \phi^{(t)} \right) = \mathcal{R}_{q}^{\text{ssr}} \left( W^{(t)}, \phi^{(t)}, b^{(t)} \right) \]

(90a)

\[
\leq \mathcal{R}_{q}^{\text{ssr}} \left( W^{(t+1)}, \phi^{(t+1)}, b^{(t)} \right)
\]

(90b)

\[
\leq \mathcal{R}_{q}^{\text{ssr}} \left( W^{(t+1)}, \phi^{(t+1)}, b^{(t)} \right)
\]

(90c)

where \( \dagger \) and \( \star \) follow from Lemmas 1 and 2, respectively. Note that for \( T_1 (t + 1) = 1 \) and \( T_2 (t + 1) = 1 \), \( W^{(t)} = W^{(t+1)} \) and \( \phi^{(t)} = \phi^{(t+1)} \). Moreover, \( T_2 (t) \geq 1 \), Lemma 2 indicates that \( \star \) holds for any alternative of \( Q_{\text{MM}} \cdot (\cdot) \) in \( A_2 (\cdot) \) which converges in a non-decreasing fashion and iterates for \( T \geq 1 \) iterations. Therefore, the inequality holds for any \( T_1 (t) = 1 \) and \( T_2 (t) \geq 1 \), and for any alternative of \( Q_{\text{MM}} \cdot (\cdot) \) with positive integer sequence of iterations.

Let us now define the expression

\[
R_k^{(t+1)} = \log \left( \frac{1 + \text{SINR}_k \left( W^{(t+1)}, \phi^{(t+1)} \right)}{1 + \text{ESNR}_k \left( W^{(t+1)}, \phi^{(t+1)} \right)} \right)
\]  

(91)

From (88), we conclude that for \( k \in [K] \) at which \( b^{(t)}_k \neq b^{(t+1)}_k \), we have

\[
\begin{cases} 
R_k^{(t+1)} > 0 & \text{if } b_k^{(t+1)} = 1 \\
R_k^{(t+1)} \leq 0 & \text{if } b_k^{(t+1)} = 0
\end{cases}
\]  

(92)
This indicates that
\[
\mathcal{R}_{q}^{\text{ssr}} \left( W^{(t+1)}, \phi^{(t+1)}, b^{(t)} \right) = \sum_{k=1}^{K} \omega_k b_k^{(t)} R_k^{(t+1)} \quad (93a)
\]
\[
\leq \sum_{k=1}^{K} \omega_k b_k^{(t+1)} R_k^{(t+1)} \quad (93b)
\]
\[
= \mathcal{R}_{q}^{\text{ssr}} \left( W^{(t+1)}, \phi^{(t+1)}, b^{(t+1)} \right) \quad (93c)
\]
\[
= \mathcal{R}_{q}^{\text{ssr}} \left( W^{(t+1)}, \phi^{(t+1)} \right) . \quad (93d)
\]

We can hence conclude the proof from (90c).

**B. Single-Loop Algorithm**

The complexity of the two-tiers algorithm can be further reduced by merging the second tier also in the main loop. This complexity reduction intuitively degrades the performance of the algorithm. However, it can boost the speed of the algorithm significantly. The numerical simulations given in the next section show that despite the existence of such a trade-off, the performance degradation in comparison to the speed gain is essentially negligible.

1) **Statement of the Algorithm:** The algorithm is represented in Algorithm 5. In this algorithm, all inner loops of the two-tiers algorithm, i.e., \( A_1 (\cdot), A_2 (\cdot) \) and the MM loop in Algorithm 1, are merged with the outer loop. In this respect, this algorithm can be observed as the special case of the two-tiers algorithm in which all the inner loops iterate only of a single iteration.

2) **Convergence Analysis:** The convergence of the single-loop algorithm is directly concluded from Theorem 2. In fact, Algorithm 5 is a special case of Algorithm 2 in which \( A_1 (\cdot), A_2 (\cdot), \) and \( Q_{\text{MM}} (\cdot) \) are run for a single iteration. Noting that Theorem 2 holds for any number of inner loop iterations, we conclude that Algorithm 5 converges in a non-decreasing fashion to its fixed-point with respect to the original objective \( \mathcal{R}_{q}^{\text{ssr}} (W, \phi) \).

**VII. NUMERICAL INVESTIGATIONS**

We investigate the performance of the proposed approach by performing several numerical experiments. A schematic view of the setting considered in simulations is shown in Fig. 2 for \( K = 4 \) legitimate UTs and \( J = 6 \) eavesdroppers. In this setting, an IRS is located at distance \( D = 25 \) m from the BS. \( K \) legitimate UTs are located uniformly and randomly around the IRS in a circle of radius \( r_{\text{IRS}} = 10 \) m. There are \( J \) eavesdroppers with uniformly generated
Algorithm 5 Single-Loop Algorithm for Precoding and Phase-Shift Tuning

**Initialization:** Set $W$ and $\phi$ to some initial values.

1: if $R_{\text{ssr}} (W, \phi)$ has not converged then 
2: Calculate $\bar{\beta}_k, \bar{t}_k, \bar{\alpha}_k$ and $\bar{w}_k$ via (54), (60a), (60b) and (57), respectively, for $k \in [K]$ 
3: Update $W = [\bar{w}_1, \ldots, \bar{w}_K]$ 
4: Calculate $\bar{f}_k, \bar{q}_k$ and $\bar{\psi}_k$ via (77b), (83a) and (83b), respectively, for $k \in [K]$ 
5: Calculate $\hat{\phi}_n = u_n^H \phi_0 + v_n - \lambda_{\text{max}} \phi_0$ with $u_n$ and $v_n$ given in (81) 
6: Update for $n \in [N]$ 
   \[ \hat{\phi}_n = -\frac{\hat{\phi}_n}{|\hat{\phi}_n|} \] 
7: Update $b_k = 1 \{\text{SINR}_k (W, \phi) > \text{ESNR}_k (W, \phi)\}$ for $k \in [K]$ 
8: Go back to line 1 
9: end if

random distances located in a circle of radius $r_{\text{BS}} = 10$ m around the BS. To calculate the noise power, a typical GSM channel with bandwidth 200 kHz has been considered. Assuming the noise spectral density to be $\log N_0 = -147$ dB/Hz, the noise power at the receiving terminals is set to $\log \sigma^2_k = \log \mu^2_j = -147$ dB.

The channel coefficients for a given UT $k$ and eavesdropper $j$ are generated according to 
\[ h_{i,k} = \varrho^L_{i,k} h^0_{i,k}, \]\[ g_{i,j} = \varrho^E_{i,j} g^0_{i,j}, \] respectively, where the index $i \in \{d, r\}$ refers to the direct and reflection paths. Here, $\varrho^L_{i,k}$ and $\varrho^E_{i,k}$ model path-loss and $h^0_{i,k}$ and $g^0_{i,j}$ take into account the impact of small-scale fading. Similarly, the channel between the BS and the IRS is written as 
\[ T = \varrho_{\text{BI}} T^0 \] where $\varrho_{\text{BI}}$ and $T^0$ model path-loss and small-scale fading effects, respectively. Throughout the simulations, we consider the standard Rayleigh fading model. This means that the entries of $h^0_{i,k}$ and $g^0_{i,j}$ for $i \in [K]$ and $j \in [J]$, as well as the entries of $T^0$ are generated i.i.d. according to a zero-mean unit-variance Gaussian distribution.
The path-loss coefficients are generated according to the following model

$$\varrho (d, \varsigma) = \frac{\varrho_{\text{ref}}}{d^\varsigma}. \tag{97}$$

In this model, $\varrho_{\text{ref}}$ denotes the path-loss at the reference distance $d = 1$ m which depends on the operating wave-length and antenna gains. Throughout the simulations, we set it to $\log \varrho_{\text{ref}} = -30$ dB. As the result, $\varrho^L_{i,k} = \varrho \left( d^L_{i,k}, \varsigma^L_{i,k} \right)$ where $d^L_{i,k}$ for $i = l$ and $i = r$ denotes the distance from legitimate UT $k$ to the BS and the distance from legitimate UT $k$ to the IRS, respectively. $\varsigma^L_{i,k}$ further denotes the path-loss exponent for the link specified by index $i$. Similarly, we set

- $\varrho^E_{i,k} = \varrho \left( d^E_{i,j}, \varsigma^E_{i,j} \right)$, where $d^E_{i,j}$ and $\varsigma^E_{i,j}$ denote the distance to the eavesdropper $j$ and its corresponding path-loss exponent for the link specified with index $i$, respectively.
- $\varrho_{\text{BI}} = \varrho (D, \varsigma_{\text{BI}})$, where $D$ and $\varsigma_{\text{BI}}$ represent the distance between the BS and the IRS and the corresponding path-loss exponent respectively.

Throughout the simulations, the distances are calculated from the realization of the randomized setting, and the path-loss exponents are set to $\varsigma^L_{d,k} = \varsigma^E_{d,k} = 3.5$ and $\varsigma^L_{r,k} = \varsigma^E_{r,k} = \varsigma_{\text{BI}} = 2.3$. The weighted secrecy sum-rate is evaluated for multiple realizations of the channel and then
averaged numerically. To indicate this point, we denote the averaged secrecy sum-rate\(^{13}\) with \(R_{\text{ssr}}\). The weights are further set to be all equal to one, i.e., \(\omega_k = 1\) for \(k \in [K]\).

A. Reference Scenarios and Benchmark

For each experiment, we evaluate the performance for the both iterative algorithms proposed in Sections VI-A and VI-B. To illustrate the performance gain obtained by employing IRSs, we consider a reference scenario in which the IRS is set off. This means that \(\beta_n = 0\) for all \(n \in [N]\). The BS in this case determines the precoding vectors via the SRZF precoding scheme proposed in [63]. We refer to this scenario as Ref. 1.

To compare the performance of the proposed algorithms with a reference point, we further consider the reference scenario Ref. 2 in which phase-shifts at the IRS are set randomly and uniformly. The linear precoding vectors are further found by applying SRZF precoding to the end-to-end equivalent channel [63]. As the state-of-the-art performance, we also compare the proposed phase-tuning algorithms with the algorithm proposed in [28]. This latter algorithm is referred to as the benchmark throughout the investigations.

We further consider an enhanced form of the setting in which the magnitudes of the signals reflected by IRS elements are also allowed to be modified. Here, we assume that the \(\beta_n\) is a tunable parameter whose value is taken from \(\beta_n \in [0, 1]\) and is updated at the same rate as the phase\(^{14}\) \(\theta_n\). As a result, tuning of the IRS reduces to a convex program which is solved tractably via a convex programming algorithm. It is worth mentioning that this is only an enhanced form of the system which we simulate for sake of comparison. In fact, due to the larger degrees of freedom, this enhanced form is expected to outperform the proposed algorithms, since in the system, amplitude modifications are not performed at the IRS. For this enhanced setting, we simulate two algorithms: The first algorithm performs alternating optimization, such that the IRS in each iteration is tuned via convex programming. This algorithm is referred to as C-Ref. 1 and can be seen as the enhanced form of the two-tiers algorithm. The second algorithm follows the single-loop algorithm with this minor modification that in each iteration the IRS tuning is performed directly via convex programming. This algorithm is referred to as C-Ref. 2 and is considered as the enhanced version of the single-loop algorithm. The convex programs in these algorithms are solved via the CVX toolbox in MATLAB [61].

\(^{13}\)This is in fact the ergodic secrecy sum-rate.

\(^{14}\)This means that the constraint \(|\phi_n| = 1\) is replaced by \(|\phi_n| \leq 1\) in the optimization problems.
B. Numerical Simulations

We start the investigations by setting $K = 4$, and $J = 6$ in the setting\textsuperscript{15}. The number of transmit antennas at the BS is set to $M = 8$, and the IRS is assumed to be equipped with $N = 128$ elements. Fig. 3 shows the weighted secrecy sum-rate against the transmit power $P_{\text{max}}$. The figure shows significant gains achieved by the proposed algorithms, when the performance is compared with the reference scenarios Ref. 1 and Ref. 2. From the figure, it is further observed that the two-tiers algorithm considerably outperform the benchmark. This is however not the case with the single-loop algorithm. Although this algorithm always outperform the benchmark, its gain becomes negligible, as the transmit power increases.

An interesting conclusion is further given by comparing the performance of convex reference algorithms C-Ref. 1 and C-Ref. 2 with the proposed algorithms. As we observe, the performance improvement achieved by replacing the unit-modulus constraints with enhanced convex ones is negligible. This is very significant for the two-tiers algorithm in which the algorithm

\textsuperscript{15}Similar to what is shown in Fig. 2.
performs almost identical in both the enhanced and the original settings\textsuperscript{16}. These tight tracks of performance in both settings confirm the efficiency of the proposed algorithms.

We now set the transmit power to $\log P_{\text{max}} = -30$ dB, and let the number of transmit antennas at the BS vary between 4 and 12. The results are plotted in Fig. 4 where the weighted secrecy sum-rate is sketched against the number of BS antennas. As the figure demonstrates, the deployment of an IRS significantly boosts the secrecy performance of the system. For instance, the secrecy sum-rate achieved in the reference scenario without an IRS, i.e., Ref. 1, with $M = 12$ transmit antennas is achieved in the IRS-aided scenario with only $M = 7$ antennas. Comparing the performance of the two tuning algorithm, it is observed from Fig. 4, that the single-loop algorithm performs slightly degraded compared to the two-tiers algorithm. This degradation is due to the inner loop merging, and is the cost we pay to reduce the computational complexity.

\textsuperscript{16}Note that for the single-loop setting, there is no guarantee that the algorithm outperforms in the enhanced setting. In fact, by merging multiple loops the algorithm may stick to a local minimum at very first iterations. This is also observed in the figure for small values of $P_{\text{max}}$ where the algorithm in the original setting outperforms slightly the enhanced setting.
As the next experiment, we consider the setting in Fig. 4 and set the number of transmit antennas to $M = 8$. We now vary the number of reflecting elements at the IRS between $N = 16$ and $N = 256$. Fig. 5 shows the secrecy sum-rate against the number of elements on the IRS. As the reference point, we have further plotted the result for Ref. 2 in which the IRS elements are tuned randomly. From the figure, one can observe that using either of the proposed algorithms, the achievable secrecy sum-rate increases linearly in terms of the number of reflecting elements with the same slope of growth. For the case with random phase-shifts, the increase in the number of IRS elements does not lead to a perceptible growth in rate.

We now investigate the resistance of the proposed algorithms against the number of eavesdroppers in the system. To this end, we keep the settings as in Fig. 5 and set the number of elements on the IRS to $N = 128$. The number of eavesdroppers is then varied from $J = 1$ to $J = 8$, and the secrecy sum-rate is plotted against $J$ in Fig. 7. For sake of comparison, we further plot the results for SRZF precoding, i.e., the reference scenarios. It is worth mentioning that from the literature, SRZF precoding is known to be highly robust against passive eavesdropping, due to the fact that it imposes explicitly the leakage suppression as a constraint in the precoder design.
Fig. 6. Weighted secrecy sum-rate against $N$ for the ideal case without phase-shift quantization at the IRS, as well as cases with quantized phase-shifts.

[63]. From Fig. 7, it is observed that the secrecy sum-rate drops for the both proposed algorithms in terms of $J$ similar to the SRZF precoding scheme. This identical behavior follows the fact that the proposed algorithms consider the secrecy rate as the objective function. This objective function implicitly constrains the information leakage in the design and leads to a behavior identical to SRZF precoding. The performance gain compared to the reference scenarios, which is observed in Fig. 7, is due to the efficient phase-shift tuning.

The proposed implementations for IRSs suggest that the phase-shifts are realized discretely by a finite resolution. This means that the phase-shifts determined by the proposed algorithms are in practice quantized with a certain number of bits. Doing so, the proposed algorithms perform degraded due to the distortion introduced after phase-shift quantization. To investigate this performance degradation, we further plot the secrecy sum-rate against the number of IRS elements for the scenario of Fig. 5 in Fig. 6 considering phase-shift quantization with $B$ bits. As the figure shows, with a resolution of $B = 4$, both algorithms perform close to the ideal case without quantization and the performance degradation is negligible.
For the last experiment, we consider the setting in Fig. 7, and set $J = 6$. The number of legitimate terminals is then increased from $K = 1$ to $K = 8$ and the secrecy sum-rate is plotted against $K$ in Fig. 8. As the figure shows, the secrecy sum-rate grows in terms of $K$. The order of growth however decreases at large choices of $K$ which is due to the higher multiuser interference in the system. Similar behavior is observed for the reference scenarios with a certain gap in the achievable rate which is due to the efficiency of phase-shift tuning in the proposed algorithms.

C. Computational Complexity

The computational complexity of both proposed algorithms shows quadratic growth in terms of the IRS dimension, i.e., $N$: In both algorithms, the dominant computational task is the update of IRS phase-shifts, i.e., $\bar{\phi} = Q_{MM}(\bar{f}, \bar{q}, \bar{\psi}, W_0)$. The computational complexity of this task for each $\phi_n$ is $O(N)$. As the update is performed entry-wise, we need $N$ iterations to determine the updated $\bar{\phi}$. This means that the computational complexity of both algorithms grows $O(N^2)$ in $N$. 

Fig. 7. Weighted secrecy sum-rate against the number of eavesdroppers $J$. 

\[ \bar{R}_{\text{sec}} \]
Fig. 8. Weighted secrecy sum-rate against the number of legitimate UTs $K$.

Fig. 9 plots the average runtime against the number of IRS elements $N$ for the proposed algorithms considering the same scenario as the one considered in Figs. 5 and 6. As it shows, despite having the same order of complexity, the single-loop algorithm runs significantly faster than the two-tiers algorithm. This follows from the fact that both inner and outer loops of the two-tiers algorithm are merged in the single-loop algorithm. To have a quantitative comparison, we further use curve-fitting to model the runtime by the quadratic polynomial function

$$Q_2(N|a,b) = aN^2 + b.$$  \hspace{1cm} (98)

The fitted curves are further plotted in the figure. From the figure, it is observed that the single-loop approach approximately reduces the computational complexity by a factor of 100. This factor is however fixed and does not scale with the system dimensions. Comparing Figs. 9 and 5, one can observe that this complexity reduction comes at the cost of performance degradation.

**VIII. CONCLUSION**

In this work, two major iterative algorithms for secure precoding and phase-shift tuning in IRS-aided MIMO systems were proposed, and their applications were investigated through multiple
Fig. 9. Runtime against the number of IRS elements $N$.

numerical experiments. The proposed algorithms determine the system parameters in terms of the CSI. This means that the proposed algorithms are required to be run, once per a coherence time interval. From the computational viewpoint, this is a tractable task in many practical scenarios.

Since the proposed algorithms consider the weighted secrecy sum-rate as the objective, the information leakage suppression is implicitly applied. This is confirmed by the numerical investigations which show similar behavior as the one observed in SRZF precoding [63]. The results further indicate that IRSs can boost the secrecy performance of the system significantly. These two findings, along with the cost-efficiency of integrating IRSs, suggest that IRS is a good candidate for establishing secure communications in MIMO systems.

The results of this study are given under several idealistic assumptions and can be extended in various aspects. A natural direction is to extend the derivations to more realistic scenarios with imperfect CSI acquisition. In this respect, one can adopt the proposed scheme to more realistic settings, e.g., by considering the study in [67]. Robustness of the proposed scheme against the unavailability of eavesdroppers’ CSI is also an interesting direction for future work. Another line of work is to develop the proposed approach for settings in which the eavesdroppers increase
their obtained leakage rate by actively contaminating the uplink pilots.

**APPENDIX A**

**PROOF OF THEOREM 1**

The proof follows two steps: In the first step, we show that for \( b \in [0, 1]^K \)

\[
\max_{\mathbf{W}, \phi} \mathcal{R}^{\text{SSF}} (\mathbf{W}, \phi) \geq \max_{\mathbf{W}, \phi, \mathbf{b}} \mathcal{R}^{\text{SSF}}_q (\mathbf{W}, \phi, \mathbf{b}).
\] (99)

To show the validity of this inequality, we define the function

\[
\mathcal{R}^{\text{SSF}}_t (\mathbf{W}, \phi, \mathbf{b}) = \sum_{k=1}^{K} \omega_k b_k \left[ \log \left( \frac{1+\text{SINR}_k (\mathbf{W}, \phi)}{1+\text{ESNR}_k (\mathbf{W}, \phi)} \right) \right]^+.
\] (100)

Since for \( x \in \mathbb{R} \log x \leq [\log x]^+ \), it is straightforward to write

\[
\mathcal{R}^{\text{SSF}}_q (\mathbf{W}, \phi, \mathbf{b}) \leq \mathcal{R}^{\text{SSF}}_t (\mathbf{W}, \phi, \mathbf{b}).
\] (101)

Now, we define the difference term

\[
\Delta = \mathcal{R}^{\text{SSF}} (\mathbf{W}, \phi) - \mathcal{R}^{\text{SSF}}_t (\mathbf{W}, \phi, \mathbf{b})
\] (102a)

\[
= \sum_{k=1}^{K} \omega_k (1 - b_k) \mathcal{R}^{\text{SSF}} (\mathbf{W}, \phi).
\] (102b)

Noting that \( \omega_k \geq 0 \) and \( b_k \in [0, 1] \), we further can write

\[
\Delta \geq 0.
\] (103)

From (101) and (103), we can write

\[
\mathcal{R}^{\text{SSF}}_q (\mathbf{W}, \phi, \mathbf{b}) \leq \mathcal{R}^{\text{SSF}}_t (\mathbf{W}, \phi, \mathbf{b}) \leq \mathcal{R}^{\text{SSF}} (\mathbf{W}, \phi),
\] (104)

which concludes that

\[
\max_{\mathbf{W}, \phi, \mathbf{b}} \mathcal{R}^{\text{SSF}}_q (\mathbf{W}, \phi, \mathbf{b}) \leq \max_{\mathbf{W}, \phi} \mathcal{R}^{\text{SSF}} (\mathbf{W}, \phi).
\] (105)

In the second step, we show that for \( b \in [0, 1]^K \)

\[
\max_{\mathbf{W}, \phi} \mathcal{R}^{\text{SSF}} (\mathbf{W}, \phi) \leq \max_{\mathbf{W}, \phi, \mathbf{b}} \mathcal{R}^{\text{SSF}}_q (\mathbf{W}, \phi, \mathbf{b}).
\] (106)

To this end, let us define the \( \mathcal{K} \subseteq [K] \) as the subset of UTs for which we have

\[
\text{SINR}_k (\mathbf{W}, \phi) \geq \text{ESNR}_k (\mathbf{W}, \phi)
\] (107)
when \( W \) and \( \phi \) are set to the optimal values. We hence have

\[
\max_{W, \phi} \mathcal{R}^{\text{ssr}}(W, \phi) = \max_{W, \phi} \sum_{k \in \mathcal{K}} w_k \mathcal{R}^a_k(W, \phi).
\] (108)

Noting that for \( k \in \mathcal{K} \),

\[
\log \left( \frac{1 + \text{SINR}_k(W, \phi)}{1 + \text{ESNR}_k(W, \phi)} \right) \geq 0,
\] (109)

we can write

\[
\mathcal{R}^a_k(W, \phi) = \log \left( \frac{1 + \text{SINR}_k(W, \phi)}{1 + \text{ESNR}_k(W, \phi)} \right)
\] (110)

and hence conclude that

\[
\max_{W, \phi} \mathcal{R}^{\text{ssr}}(W, \phi) = \max_{W, \phi} \mathcal{R}^{\text{ssr}}(W, \phi, b_{/\mathcal{K}})
\] (111)

where \( b_{/\mathcal{K}} = [b_{/\mathcal{K}1}, \ldots, b_{/\mathcal{K}K}]^T \) with

\[
b_{/\mathcal{K}k} = \begin{cases} 
1 & k \in \mathcal{K} \\
0 & k \notin \mathcal{K}
\end{cases}
\] (112)

Noting that \( b_{/\mathcal{K}k} \in [0, 1]^K \), we have

\[
\max_{W, \phi, b} \mathcal{R}^{\text{ssr}}(W, \phi, b) \geq \max_{W, \phi} \mathcal{R}^{\text{ssr}}(W, \phi, b_{/\mathcal{K}}).
\] (113)

This concludes (106). Considering the inequalities in (99) and (106), Theorem 1 is proved.

**APPENDIX B**

**DERIVATION OF THE MM-BASED ALGORITHM**

The MM method approximates the solution with the limit of a sequence of feasible points. This sequence is derived by optimizing sequentially an objective function which majorizes\(^{17}\) the original objective at the feasible point derived in the previous optimization. We illustrate the method through the derivations. More details on the MM method can be followed in [58], [66], [68]–[70].

We start the derivations by rewriting the optimization problem \( \hat{Q}_2^B \). Let \( Q_2(\phi) \) denote the negative objective of \( \hat{Q}_2^B \), i.e.,

\[
Q_2(\phi) = -Q_2^m(\phi, \bar{f}) - \frac{\omega_k b_{0k} \left( 1 + \bar{\psi}_k \right)}{B_{2Uk}^e} A_{2k}^e(\phi).
\] (114)

\(^{17}\)The concept of majorization is defined in the sequel.
By standard derivations, one can rewrite $Q_2(\phi)$ as

$$Q_2(\phi) = \phi^H U \phi + 2 \Re \{ \phi^H v \} + C$$  \hspace{1cm} (115)$$

for some $C$ which is fixed in $\phi$ and $U$ and $v$ as defined in Algorithm 1. As the result, $\hat{Q}_2^B$ is rewritten as

$$\min_{\phi} \phi^H U \phi + 2 \Re \{ \phi^H v \}$$  \hspace{1cm} (116a)$$

subject to $|\phi_n| = 1, \forall n \in [N]$.  \hspace{1cm} (116b)$$

In order to address this problem via the MM method, we need to follow two steps:

1) Define a sequence of majorizations for the objective in (116).

2) Determine the sequence of the solutions to the majorizations series.

Before we start with the first step, let us define the concept of majorization.

**Definition 3** (Majorization). Consider $f(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$. The function $m(\cdot|x_0) : \hat{\mathcal{X}} \rightarrow \mathbb{R}$ with $x_0 \in \mathcal{X}$ and $\mathcal{X} \subseteq \hat{\mathcal{X}}$ majorizes $f(\cdot)$ at $x_0$, if the following constraints are satisfied:

$$f(x) \leq m(x|x_0), \forall x \in \mathcal{X},$$  \hspace{1cm} (117a)$$

$$f(x_0) = m(x_0|x_0).$$  \hspace{1cm} (117b)$$

To find majorization functions of the objective in (116), we invoke the following lemma whose proof can be followed in [66, Lemma 1].

**Lemma 3** (Lemma 1 of [66]). Let $U \in \mathbb{C}^{N \times N}$ and $M \in \mathbb{C}^{N \times N}$ be Hermitian matrices satisfying $M \succeq U$. At any point $x_0 \in \mathbb{C}^N$, the quadratic function $f(x) = x^H U x$ is majorized by

$$\Omega(x|x_0) = x^H M x + 2 \Re \{ x^H (U - M) x_0 \} + x_0^H (M - U) x_0$$  \hspace{1cm} (118)$$

Let $M = \lambda_{\text{max}} I_N$ in Lemma 3, where $\lambda_{\text{max}}$ denotes the maximum eigenvalue of $U$. Hence, at a feasible point $\phi_0$, we can write

$$\phi^H U \phi + 2 \Re \{ \phi^H v \} \leq \hat{\Omega}(\phi|\phi_0),$$  \hspace{1cm} (119)$$

where $\hat{\Omega}(\phi|\phi_0)$ is given by

$$\hat{\Omega}(\phi|\phi_0) = \lambda_{\text{max}} (\|\phi\|^2 + \|\phi_0\|^2) + 2 \Re \{ \phi^H (U \phi_0 + v - \lambda_{\text{max}} \phi_0) \} - \phi_0^H U \phi_0.$$  \hspace{1cm} (120)$$
Noting that for any feasible point \( \phi \), we have \( \| \phi \|^2 = N \), we can further write

\[
\hat{\Omega} (\phi | \phi_0) = 2N\lambda_{\max} + 2 \Re \{ \phi^H (U\phi_0 + v - \lambda_{\max}\phi_0) \} - \phi_0^H U\phi_0. 
\] (121)

Given the majorization function \( \hat{\Omega} (\phi | \phi_0) \), the MM method starts from a feasible point \( \phi^{(0)} \) and constructs the sequence \( \{\phi^{(t)}\} \) for \( t \in \mathbb{Z}^+ \) iteratively as

\[
\phi^{(t+1)} = \arg\min_{\phi} \hat{\Omega} (\phi | \phi^{(t)}) 
\]

subject to \( |\phi_n| = 1, \; \forall n \in [N] \). (122b)

The limit of this sequence, as the algorithm converges, is considered to be the solution. Unlike the original problem, the solution to (122) is easily given by

\[
\phi_n^{(t+1)} = -\frac{u_n^H \phi^{(t)} + v_n - \lambda_{\max}\phi_n^{(t)}}{|u_n^H \phi^{(t)} + v_n - \lambda_{\max}\phi_n^{(t)}|} 
\] (123)

with \( u_n^H \) denoting the \( n \)-th row of the matrix \( U \). This concludes the derivation of Algorithm 1.

**APPENDIX C**

**A BCD-TYPE PHASE-SHIFT UPDATE ALGORITHM**

In this appendix, we tackle the non-convex unit modulus optimization \( \hat{Q}_2^B \) via the BCD method. Starting from an initial vector of phase-shifts, i.e., \( \phi^{(0)} \), we alternately update each phase-shift in iteration \( t \), i.e., each entry of \( \phi^{(t)} \), by marginally optimizing it while treating the other phase-shifts as fixed variables whose values are calculated in iteration \( t - 1 \). This means that in the \( t \)-th iteration, \( \phi_n^{(t)} \) is found as

\[
\phi_n^{(t)} = \arg\min_{|\varphi|=1} Q_2 (x_{n}^{(t)} (\varphi))
\] (124)

where \( Q_2 (\cdot) \) is given in (115), and \( x_{n}^{(t)} (\cdot) \) is defined as

\[
x_{n}^{(t)} (\varphi) = \left[ \phi_1^{(t-1)}, \ldots, \phi_n^{(t-1)} \varphi, \phi_{n+1}^{(t-1)}, \ldots, \phi_N^{(t-1)} \right]^T
\] (125)

with \( \phi_n^{(t-1)} \) denoting the \( n \)-th entry of \( \phi^{(t-1)} \).

Defining the function \( f^{(t)} (\varphi) = Q_2 (x_{n}^{(t)} (\varphi)) \), we can use the Hermitian symmetry of \( U \) to show that

\[
f^{(t)} (\varphi) = |U|_{nn} |\varphi|^2 + 2 \Re \{ a_{n}^{(t)} \varphi^* \} + c_{n}^{(t)}
\] (126)
where \( a_1^{(t)} \) and \( a_2^{(t)} \) are given by

\[
a_n^{(t)} = v_n + \tilde{u}_n^H \tilde{\phi}^{(t-1)}
\]

with \( \tilde{u}_n, \tilde{\phi}^{(t-1)} \in \mathbb{C}^{N-1} \) being constructed from \( u_n \) and \( \phi^{(t-1)} \) by excluding their \( n \)-th entries, respectively, and \( a_n^{(t)} \) is a constant in terms of \( \varphi \). Noting that \( |\varphi| = 1 \), (124) reduces to

\[
\phi_n^{(t)} = \arg\min_{|\varphi|=1} \Re \{ a_n^{(t)} \varphi^* \}
\]

whose solution is given in a closed form as

\[
\phi_n^{(t)} = -\frac{a_n^{(t)}}{|a_n^{(t)}|}.
\]

After multiple iterations, the solution of \( \hat{Q}_2^B \) is approximated with the converged phase-shifts. The final algorithm is summarized in Algorithm 6.

**Algorithm 6 Phase-Shift Update via the BCD Method**

**Input:** \( f, \bar{q}, \psi \) and \( W_0 \)

**Initialization:** Set a feasible initial point \( \phi = \phi^{(0)} \), and define

\[
\mathcal{R}_{SSR}^M (\phi) = Q_2^m (\phi, \bar{f}) + \frac{\omega_k \tilde{b}_{0k} (1 + \tilde{\psi}_k)}{B_{2\tilde{U}k}} A_{2k}^e (\phi)
\]

1: if \( \mathcal{R}_{SSR}^M (\phi) \) has not converged then
2: Calculate \( U \) and \( v \) from (81)
3: Update \( a_n^{(t)} \) via (127) for \( n \in [N] \)
4: Update \( \phi_n^{(t)} \) via (129) for \( n \in [N] \)
5: Set \( \phi = \phi^{(t)} \), \( t \leftarrow t + 1 \) and go back to line 1
6: end if

Algorithm 6 is compared with Algorithm 1 by performing a simple numerical experiment. This experiment considers a setting consistent to the one presented in Fig. 2. In this setting, we set the number of eavesdroppers to \( J = 6 \), the number of legitimate UTs to \( K = 4 \), the number of BS antennas to \( M = 8 \) and the number of IRS elements to \( N = 16 \). The transmit power is further set to \( \log P_{max} = 0 \) dB and the channel parameters are set as described in Section VII-B. For this setting, we apply the two-tiers algorithm, i.e., Algorithm 2 two times: once the inner loop, i.e., Algorithm 4, is performed via the MM algorithm and once with Algorithm 6. Fig. 10
Fig. 10. Comparing the BCD-based phase tuning approach to the one which performs the MM algorithm.

shows the achievable secrecy sum-rate against the number of iterations in the main loop of Algorithm 2. As the figure shows, the algorithms converge to the target rate almost identical with respect to the number of iterations. To compare the computational complexity of these algorithms, we further plot the achievable secrecy sum-rate against the runtime for these two algorithms. The result is shown in Fig. 11. As the figure shows, using Algorithm 6 leads to a slightly faster convergence compared to the MM algorithm.

**APPENDIX D**

**PROOF OF LEMMA 1**

Following the notation of Algorithm 3, let $\phi_0 = \tilde{\phi}$ and $b_0 = \tilde{b}$ denote the vector of phase-shifts and auxiliary variables fixed as inputs of the algorithm, and let $W$ and $W_{new}$ represent the precoding matrix updated at the end of iteration $t$ and $t + 1$, respectively. We define the
Fig. 11. Comparing the complexity of the BCD-based and MM-based phase tuning approaches.

The following functions:

\[ f_Q(\beta, \gamma, W) = Q_{1}^m(W, \beta) + Q_{1}^e(W, \gamma) + \log B_{1U} \]

\[ R_{ssr}^M(W) = R_{ssr}^q(W, \phi, b) \]

with \( Q_{1}^m(W, \beta) \), \( Q_{1}^e(W, \gamma) \), and \( B_{1U} \) as defined in Section V-A. Moreover, let us define vectors

\[ \beta^F(W) = [\beta_1^F(W), \ldots, \beta_K^F(W)] \]

\[ \gamma^F(W) = [\gamma_1^F(W), \ldots, \gamma_K^F(W)] \]

for a given precoding matrix \( W \), where

\[ \beta_k^F(W) = \sqrt{\frac{\omega_k b_k \left( 1 + \frac{A_{1k}^m(W)}{B_{1k}^m(W)} \right) h_k^H(\phi) w_k}{A_{1k}^m(W) + B_{1k}^m(W)}} \]

\[ \gamma_k^F(W) = \sqrt{\frac{\omega_k b_k \left( 1 + \frac{A_{1k}^e(W)}{B_{1k}^e(W)} \right)}{B_{1U}^e} \sqrt{A_{1k}^e(W)}} \]
By substitution, it is straightforward to show that in iteration $t$ of Algorithm 3, $\bar{\beta} = \beta^F(\bar{W})$ and $\bar{\gamma} = \gamma^F(\bar{W})$ and that

$$R_{ssr}^{\text{M}}(W) = f_Q(\beta^F(\bar{W}), \gamma^F(\bar{W}), \bar{W}).$$

(133)

To start the proof, we note that for a fixed $W$, the marginal function $f_Q(\beta, \gamma, W)$ is a concave function in terms of $\beta$ and $\gamma$ whose maximum is at $\beta = \beta^F(W)$ and $\gamma = \gamma^F(W)$. As a result, for a given $W$, we have

$$f_Q(\bar{\beta}, \bar{\gamma}, W) \leq f_Q(\bar{\beta}, \bar{\gamma}, W_{\text{new}}).$$

(134a)

$$= R_{ssr}^{\text{M}}(W).$$

(134b)

Furthermore, the marginal function $f_Q(\bar{\beta}, \bar{\gamma}, W)$ is a concave function in $W$. The quadratic problem $Q_1^B$ indicates that the maximum of this marginal function is at $W = W_{\text{new}}$, when we set $\beta = \bar{\beta}$ and $\gamma = \bar{\gamma}$. Thus, for a given $W$, we have

$$f_Q(\bar{\beta}, \bar{\gamma}, W) \leq f_Q(\bar{\beta}, \bar{\gamma}, W_{\text{new}}).$$

(135)

Considering the above inequalities, we can write

$$R_{ssr}^{\text{q}}(W, \bar{\phi}, \bar{b}) = R_{ssr}^{\text{M}}(W)$$

$$\overset{\dagger}{=} f_Q(\bar{\beta}, \bar{\gamma}, W)$$

$$\overset{\star}{=} f_Q(\bar{\beta}, \bar{\gamma}, W_{\text{new}})$$

$$\overset{\clubsuit}{=} f_Q(\beta^F(W_{\text{new}}), \gamma^F(W_{\text{new}}), W_{\text{new}})$$

$$= R_{ssr}^{\text{M}}(W_{\text{new}})$$

$$= R_{ssr}^{\text{q}}(W_{\text{new}}, \bar{\phi}, \bar{b})$$

(136a)

(136b)

(136c)

(136d)

(136e)

(136f)

where $\dagger$ follows (133), $\star$ comes from (135), and $\clubsuit$ is concluded by (134b).

**APPENDIX E**

**PROOF OF LEMMA 2**

The proof follows the same approach as in Appendix D. Assuming the initialization $W_0 = \bar{W}$ and $b_0 = \bar{b}$, we start the proof by defining

$$f_Q(f, \bar{\phi}, \bar{b}) = Q_2^m(\phi, f) + Q_2^e(\phi, \bar{\phi}) + \sum_{k=1}^{K} \omega_k \log B_{2Uk}^{e},$$

(137a)

$$R_{ssr}^{\text{M}}(\phi) = R_{ssr}^{\text{q}}(W, \phi, \bar{\phi})$$

(137b)
with $Q_{2}^m (\phi, f)$, $Q_{2}^e (\phi, \varpi)$ and $B_{2Uk}^e$ defined in Section V-B. We further define

$$f^F(\phi) = \left[ f_{1}^F(\phi), \ldots, f_{K}^F(\phi) \right],$$  \hspace{1cm} (138a)

$$\varpi^F(\phi) = \left[ \varpi_{1}^F(\phi), \ldots, \varpi_{K}^F(\phi) \right]$$  \hspace{1cm} (138b)

for a given $\phi$, where

$$f_{k}^F(\phi) = \sqrt{\omega_{k} b_{k} \left( 1 + \frac{A_{mk}^m(\phi)}{B_{2k}^m(\phi)} \right) \left( h_{d,k}^H \bar{w}_{k} + \phi^H H_{k}^H \bar{w}_{k} \right) \| h_{d,k}^H \bar{W} + \phi^H H_{k}^H \bar{W} \|^2 + \sigma_{k}^2},$$  \hspace{1cm} (139a)

$$\varpi_{k}^F(\phi) = \sqrt{\omega_{k} b_{k} \left( 1 + \frac{A_{ek}^e(\phi)}{B_{2k}^e(\phi)} \right) \frac{B_{2Uk}^e}{A_{2k}^e(\phi)}}. $$  \hspace{1cm} (139b)

By substitution, it is easily shown that for a given $\phi$

$$R_{M}^{ssr}(\phi) = f_Q\left( f^F(\phi), \varpi^F(\phi), \phi \right).$$  \hspace{1cm} (140)

Let $\bar{\phi}$ and $\phi_{new}$ indicate the phase-shift vectors updated at the end of iteration $t$ and $t + 1$, respectively. The convergence of the MM algorithm, i.e., Algorithm 1, guarantees

$$f_Q\left( f^F(\bar{\phi}), \varpi^F(\bar{\phi}), \bar{\phi} \right) \leq f_Q\left( f^F(\bar{\phi}), \varpi^F(\bar{\phi}), \phi_{new} \right).$$  \hspace{1cm} (141)

This inequality holds as an identity, if Algorithm 1 iterates only for one iteration. Following concavity of the marginal function $f_Q(\mathbf{f}, \varpi, \phi)$ in terms of $\mathbf{f}$ and $\varpi$ with maximum being at $f^F(\phi)$ and $\varpi^F(\phi)$, for a given $\phi$, we can write that

$$f_Q\left( f^F(\bar{\phi}), \varpi^F(\bar{\phi}), \phi \right) \leq f_Q\left( f^F(\phi), \varpi^F(\phi), \phi \right)$$  \hspace{1cm} (142)

for any feasible $\phi$.

We hence can use the above inequalities and write

$$R_{q}^{ssr}(\bar{W}, \bar{\phi}, \bar{b}) = R_{M}^{ssr}(\bar{\phi})$$  \hspace{1cm} (143a)

$$= f_Q\left( f^F(\bar{\phi}), \varpi^F(\bar{\phi}), \bar{\phi} \right)$$  \hspace{1cm} (143b)

$$\leq f_Q\left( f^F(\bar{\phi}), \varpi^F(\bar{\phi}), \phi_{new} \right)$$  \hspace{1cm} (143c)

$$\leq f_Q\left( f^F(\phi_{new}), \varpi^F(\phi_{new}), \phi_{new} \right)$$  \hspace{1cm} (143d)

$$= R_{M}^{ssr}(\phi_{new})$$  \hspace{1cm} (143e)

$$= R_{q}^{ssr}(\bar{W}, \phi_{new}, \bar{b}).$$  \hspace{1cm} (143f)
Noting that (141) is the only constraint we need for the convergence proof, in terms of phase-tuning update algorithm, we conclude that the proof is valid for any alternative of Algorithm 1 which converges to its fixed-point in a non-decreasing fashion.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” IEEE Network, vol. 34, no. 3, pp. 134–142, Jun. 2020.
[2] E. Yaacoub and M.-S. Alouini, “A key 6G challenge and opportunity—Connecting the base of the pyramid: A survey on rural connectivity,” Proc. IEEE, vol. 108, no. 4, pp. 533–582, Apr. 2020.
[3] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, “The roadmap to 6G: AI empowered wireless networks,” IEEE Commun. Mag., vol. 57, no. 8, pp. 84–90, Aug. 2019.
[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Commun. Mag., vol. 52, no. 2, pp. 186–195, Feb. 2014.
[5] J. Hoydis, S. Ten Brink, and M. Debbah, “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?” IEEE J. sel. Areas Commun., vol. 31, no. 2, pp. 160–171, Jan. 2013.
[6] P. Yang, Y. Xiao, M. Xiao, and S. Li, “6G wireless communications: Vision and potential techniques,” IEEE Network, vol. 33, no. 4, pp. 70–75, Jul. 2019.
[7] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Armon, and G. Chen, “Millimeter wave communication: A comprehensive survey,” IEEE Commun. Surv. & Tut., vol. 20, no. 3, pp. 1616–1653, Jun. 2018.
[8] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!” IEEE Access, vol. 1, pp. 335–349, May 2013.
[9] C.-X. Wang, J. Huang, H. Wang, X. Gao, X. You, and Y. Hao, “6G wireless channel measurements and models: Trends and challenges,” IEEE Veh. Techn. Mag., vol. 15, no. 4, pp. 22–32, Dec. 2020.
[10] J. An, K. Yang, J. Wu, N. Ye, S. Guo, and Z. Liao, “Achieving sustainable ultra-dense heterogeneous networks for 5G,” IEEE Commun. Mag., vol. 55, no. 12, pp. 84–90, Dec. 2017.
[11] C. Niu, Y. Li, R. Q. Hu, and F. Ye, “Fast and efficient radio resource allocation in dynamic ultra-dense heterogeneous networks,” IEEE Access, vol. 5, pp. 1911–1924, Feb. 2017.
[12] S. Zhang, J. Liu, H. Guo, M. Qi, and N. Kato, “Envisioning device-to-device communications in 6G,” IEEE Network, vol. 34, no. 3, pp. 86–91, Jun. 2020.
[13] S. Malkowsky, J. Vieira, L. Liu, P. Harris, K. Nieman, N. Kundargi, I. C. Wong, F. Tufvesson, V. Öwall, and O. Edfors, “The world’s first real-time testbed for massive MIMO: Design, implementation, and validation,” IEEE Access, vol. 5, pp. 9073–9088, May 2017.
[14] Y. Huo, X. Dong, W. Xu, and M. Yuen, “Enabling multi-functional 5G and beyond user equipment: A survey and tutorial,” IEEE Access, vol. 7, pp. 116 975–117 008, Aug. 2019.
[15] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, “Intelligent reflecting surface aided wireless communications: A tutorial,” arXiv preprint arXiv:2007.02759, Jul. 2020.
[16] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, “Multicell MIMO communications relying on intelligent reflecting surfaces,” IEEE Trans. on Wireless Commun., vol. 19, no. 8, pp. 5218–5233, Aug. 2020.
[17] Ö. Özdogan, E. Björnson, and E. G. Larsson, “Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling,” IEEE Wireless Commun. Letters, vol. 9, no. 5, pp. 581–585, Dec. 2019.
[18] E. Björnson, Ö. Özdogan, and E. G. Larsson, “Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?” IEEE Wireless Communications Letters, vol. 9, no. 2, pp. 244–248, Feb. 2019.

[19] S. Hu, F. Rusek, and O. Edfors, “Beyond massive MIMO: The potential of data transmission with large intelligent surfaces,” IEEE Trans. Signal Process., vol. 66, no. 10, pp. 2746–2758, May 2018.

[20] E. Björnson, L. Sanguinetti, H. Wyneersch, J. Hoydis, and T. L. Marzetta, “Massive MIMO is a reality—What is next?: Five promising research directions for antenna arrays,” Digital Signal Process., vol. 94, pp. 3–20, Nov. 2019.

[21] 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.

[22] E. Björnson and L. Sanguinetti, “Power scaling laws and near-field behaviors of massive MIMO and intelligent reflecting surfaces,” IEEE Open J. Commun. Soc., vol. 1, pp. 1306–1324, Sep. 2020.

[23] D.-W. Yue, H. H. Nguyen, and Y. Sun, “Analysis of intelligent reflecting surface-assisted mmWave doubly massive-MIMO communications,” arXiv preprint arXiv:2003.00282, Mar. 2020.

[24] V. Jamali, A. M. Tulino, G. Fischer, R. Müller, and R. Schober, “Intelligent reflecting and transmitting surface aided millimeter wave massive MIMO,” arXiv preprint arXiv:1902.07670, 2019.

[25] A. Bereyhi, V. Jamali, R. R. Müller, G. Fischer, R. Schober, and A. M. Tulino, “PAPR-limited precoding in massive MIMO systems with reflect-and-transmit-array antennas,” in Proc. 53rd Asilomar Conf. on Signals, Systems, and Computers. IEEE, Pacific Grove, CA, USA, Nov. 2019, pp. 1690–1694.

[26] A. Bereyhi, V. Jamali, R. R. Müller, A. M. Tulino, G. Fischer, and R. Schober, “A single-RF architecture for multiuser massive MIMO via reflecting surfaces,” in Proc. IEEE International Conf. on Acoustics, Speech and Signal Process. IEEE, Barcelona, Spain, May 2020, pp. 8688–8692.

[27] R. Karasik, O. Simeone, M. Di Renzo, and S. Shamai, “Single-RF multi-user communication through reconfigurable intelligent surfaces: An information-theoretic analysis,” arXiv preprint arXiv:2101.07556, Jan. 2021.

[28] L. Dong and H.-M. Wang, “Enhancing secure MIMO transmission via intelligent reflecting surface,” IEEE Trans. Wireless Commun., vol. 19, no. 11, pp. 7543–7556, Nov. 2020.

[29] X. Lu, W. Yang, X. Guan, Q. Wu, and Y. Cai, “Robust and secure beamforming for intelligent reflecting surface aided mmWave MISO systems,” IEEE Wireless Communications Letters, vol. 9, no. 12, pp. 2068–2072, Dec. 2020.

[30] S. Hong, C. Pan, H. Ren, K. Wang, and A. Nallanathan, “Artificial-noise-aided secure MIMO wireless communications via intelligent reflecting surface,” IEEE Transactions on Communications, vol. 68, no. 12, pp. 7851–7866, Dec. 2020.

[31] Y. Cao and T. Lv, “Intelligent reflecting surface aided multi-user millimeter-wave communications for coverage enhancement,” arXiv preprint arXiv:1910.02398, 2019.

[32] 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.

[33] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, Z. Qin, M. Di Renzo, and N. Al-Dhahir, “Reconfigurable intelligent surfaces: Principles and opportunities,” arXiv preprint arXiv:2007.03435, 2020.

[34] Q. Wu and R. Zhang, “Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network,” IEEE Communications Magazine, vol. 58, no. 1, pp. 106–112, Jan. 2019.

[35] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, “Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications,” IEEE Commun. Letters, vol. 23, no. 9, pp. 1488–1492, Sep. 2019.

[36] M. Cui, G. Zhang, and R. Zhang, “Secure wireless communication via intelligent reflecting surface,” IEEE Wireless Commun. Letters, vol. 8, no. 5, pp. 1410–1414, May 2019.

[37] H.-M. Wang, J. Bai, and L. Dong, “Intelligent reflecting surfaces assisted secure transmission without eavesdropper’s CSI,” IEEE Signal Process. Letters, vol. 27, pp. 1300–1304, Jul. 2020.
[38] L. Yang, J. Yang, W. Xie, M. O. Hasna, T. Tsiftsis, and M. Di Renzo, “Secrecy performance analysis of RIS-aided wireless communication systems,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12 296–12 300, Oct. 2020.

[39] Y. Song, M. R. Khandaker, F. Tariq, and K.-K. Wong, “Truly intelligent reflecting surface-aided secure communication using deep learning,” *arXiv preprint arXiv:2004.03056*, 2020.

[40] X. Guan, Q. Wu, and R. Zhang, “Intelligent reflecting surface assisted secrecy communication: Is artificial noise helpful or not?” *IEEE Wireless Commun. Letters*, vol. 9, no. 6, pp. 778–782, Sep. 2020.

[41] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, “Resource allocation for secure IRS-assisted multiuser MISO systems,” in *2019 IEEE Globecom Workshops*, Waikoloa, HI, USA, Dec., 2019, pp. 1–6.

[42] A. Bereyhi, S. Asaad, R. R. Müller, R. F. Schaefer, and H. V. Poor, “Secure transmission in IRS-assisted MIMO systems with active eavesdroppers,” *arXiv preprint arXiv:2010.07989*, 2020.

[43] A. Bereyhi, S. Asaad, R. R. Müller, R. F. Schaefer, and A. M. Rabiei, “On robustness of massive MIMO systems against passive eavesdropping under antenna selection,” in *Proc. of IEEE Global Commun. Conf.*, Abu Dhabi, UAE, Dec., 2018, pp. 1–7.

[44] A. Bereyhi, S. Asaad, R. R. Müller, R. F. Schaefer, G. Fischer, and H. Vincent Poor, “Securing massive MIMO systems: Secrecy for free with low-complexity architectures,” *IEEE Trans. Wireless Commun.*, vol. Early Access, pp. 1–15, Apr. 2021.

[45] 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.

[46] Q. Wang, F. Zhou, R. Q. Hu, and Y. Qian, “Energy-efficient beamforming and cooperative jamming in IRS-assisted MISO networks,” in *IEEE Int. Conf. Commun.*, Jun. Virtual Conf., 2020, pp. 1–7.

[47] 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. 82 599–82 612, 2019.

[48] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, “Robust and secure wireless communications via intelligent reflecting surfaces,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov., 2020.

[49] 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.

[50] ———, “Fractional programming for communication systems – part II: Uplink scheduling via matching,” *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, May 2018.

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

[52] Z. Wang, L. Liu, and S. Cui, “Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis,” *arXiv preprint arXiv:1912.11783*, Apr. 2020.

[53] Z.-Q. He and X. Yuan, “Cascaded channel estimation for large intelligent metasurface assisted massive MIMO,” *IEEE Wireless Commun. Letters*, vol. 9, no. 2, pp. 210–214, Feb. 2020.

[54] R. Liu, T. Liu, H. V. Poor, and S. Shamai, “Multiple-input multiple-output Gaussian broadcast channels with confidential messages,” *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4215–4227, Sep. 2010.

[55] ———, “New results on multiple-input multiple-output broadcast channels with confidential messages,” *IEEE Trans Inf. Theory*, vol. 59, no. 3, pp. 1346–1359, Mar. 2013.

[56] V. S. Asadchy, M. Alboooyeh, S. N. Tcvetkova, A. Diaz-Rubio, Y. Ra’di, and S. A. Tretyakov, “Perfect control of reflection and refraction using spatially dispersive metasurfaces,” *Phys. Rev. B*, vol. 94, p. 075142, Aug. 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.94.075142
[57] 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.

[58] 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.

[59] D. P. a. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific Belmont, MA, USA: Athena Scientific, 1999.

[60] Q. Shi and M. Hong, “Spectral efficiency optimization for millimeter wave multiuser MIMO systems,” *IEEE J. Sel. Topics in Signal Process.*, vol. 12, no. 3, pp. 455–468, Jun. 2018.

[61] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” http://cvxr.com/cvx, Mar. 2014.

[62] ———, “Graph implementations for nonsmooth convex programs,” in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[63] S. Asaad, A. Bereyhi, R. R. Müller, and R. F. Schaefer, “Secure regularized zero forcing for multiuser MIMOME channels,” IEEE, pp. 1108–1113, Pacific Grove, CA, USA, Nov., 2019.

[64] W.-K. K. Ma, “Semidefinite relaxation of quadratic optimization problems and applications,” *IEEE Signal Process. Mag.*, vol. 1053, no. 5888/10, 2010.

[65] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[66] J. Song, P. Babu, and D. P. Palomar, “Optimization methods for designing sequences with low autocorrelation sidelobes,” *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3998–4009, Aug. 2015.

[67] L. You, J. Xiong, Y. Huang, D. W. K. Ng, C. Pan, W. Wang, and X. Gao, “Reconfigurable intelligent surfaces-assisted multiuser MIMO uplink transmission with partial CSI,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5613–5627, Sep. 2021.

[68] D. R. Hunter and K. Lange, “A tutorial on MM algorithms,” *The American Statistician*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[69] P. Stoica and Y. Selen, “Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: A refresher,” *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 112–114, Jan. 2004.

[70] M. Razaviyayn, M. Hong, and Z.-Q. Luo, “A unified convergence analysis of block successive minimization methods for nonsmooth optimization,” *SIAM J. Opt.*, vol. 23, no. 2, pp. 1126–1153, Jun. 2013.