On Linear Precoding for the Two-User MISO Broadcast Channel with Confidential Messages and Per-Antenna Constraints

Ayman Mostafa, *Member, IEEE,* and Lutz Lampe, *Senior Member, IEEE*

***Abstract*—We study the design of linear precoders for secure transmission in the two-user multiple-input single-output (MISO) broadcast channel with confidential messages (BC-CM). The transmitter has multiple antennas, and each user has a single receive antenna. Two independent messages are simultaneously transmitted, one intended for each user, and each message should be kept confidential from the other user. Assuming real-valued transmitted signals, we design the linear precoders subject to total and per-antenna average power constraints, and also subject to amplitude constraints. In both cases, we tackle the design problem via weighted secrecy sum rate maximization. The resulting problem, however, involves a fractional objective, making it nonconvex and difficult to solve. Nevertheless, we show that this difficult problem can be transformed into a more tractable problem, for which a solution can be obtained by an iterative search algorithm. In addition, we characterize a condition under which the obtained solution is guaranteed to be optimal. Furthermore, we show that the problem formulation and solution approach can be easily extended to handle the robust version of the design problem with uncertain channel information. We provide numerical examples to demonstrate the performance of the proposed precoder in terms of the achievable secrecy rate regions subject to the aforementioned constraints. We also demonstrate the performance of the robust precoder under different channel uncertainty levels.**

***Index Terms*—Amplitude constraint, MISO broadcast channel with confidential messages, per-antenna power constraint, robust linear precoding, secrecy rate region.**

I. Introduction

THE foundations of information-theoretic security were laid down by Wyner in his seminal paper [1] that studied the problem of secret communication over the degraded broadcast channel. In that paper, Wyner introduced the so-called *wiretap channel* model to describe the scenario in which the transmitter has a secret message intended for one receiver, while the other receiver, whose channel is degraded, acts only as an eavesdropper. Wyner also proposed the notion of *secrecy capacity* as a performance measure that specifies the maximum communication rate that guarantees reliable reception of the secret message by the intended receiver, and entire hiddenness from the eavesdropper. This motivated the characterization of the secrecy capacity of the scalar Gaussian wiretap channel [2]. Wyner's model was then extended to the (nondegraded) broadcast channel in which the eavesdropper's channel need not be degraded [3]. Such an extension has ultimately led to the characterization of the secrecy capacity of the multiple-input single-output (MISO) and multiple-input multiple-output (MIMO) Gaussian wiretap channels [4], [5].

The wiretap channel model was further extended in [6] to the two-user broadcast channel with confidential messages (BC-CM), and the secrecy capacity region was adopted as the performance measure. This model captures the practically relevant scenario in which the transmitter has two independent secret messages, one intended for each receiver, and each message should be kept confidential from the other receiver. Achievability of the secrecy capacity region of the two-user MISO BC-CM was established in [7] using the so-called *secret dirty-paper coding* (S-DPC) scheme under the total (average) power constraint. This coding scheme was then extended in [8] to the MIMO BC-CM, and it was shown that the secrecy capacity region is rectangular under the matrix power (or input covariance matrix) constraint. Under the total power constraint, however, the secrecy capacity region can be only found by performing an exhaustive search over the set of all input covariance matrices that satisfy the total power constraint. Due to the complexity of S-DPC and the lack of a simple solution to the practical case of total power constraint, the authors in [9] proposed a low-complexity linear precoding scheme for the two-user MIMO BC-CM based on generalized singular value decomposition. The work in [10] also characterized a secrecy rate region for the two-user MIMO BC-CM under the total power constraint via formulating a nonconvex weighted secrecy sum rate maximization problem. An iterative algorithm based on a block successive lower-bound maximization method was proposed to solve such a nonconvex problem.

In practical multiple-antenna systems, each antenna element is equipped with a separate power amplifier. Therefore, per-antenna power constraints are often necessary to model hardware limitations in practical systems. With such constraints, however, the design problem may become more difficult to handle. Nonetheless, several works in the literature have considered the transmitter optimization problem subject to per-antenna power constraints [11]–[15]. In [11], the authors considered the design of zero-forcing (ZF) linear precoders for weighted sum rate maximization in the multi-user MISO broadcast channel. For the multi-user MIMO case, the authors

Manuscript received January 9, 2017; revised June 15, 2017 and July 28, 2017; accepted August 5, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yue Rong. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

A. Mostafa and L. Lampe are with the Department of Electrical and Computer Engineering, The University of British Colombia, Vancouver, BC, V6T 1Z4, Canada (emails: amostafa@ece.ubc.ca, lampe@ece.ubc.ca).

in [12] studied the design of linear transceivers for weighted sum rate and minimum-rate maximization. More recently, the authors in [13] have shown that beamforming is optimal for the point-to-point MISO channel subject to joint total and per-antenna power constraints. Furthermore, the capacity region of the bidirectional MISO channel was characterized in [14] under per-antenna power constraints. For the MIMO wiretap channel, the problem of finding the optimal transmit covariance matrix appears to be difficult to solve, even under the total power constraint. Therefore, the authors in [15] have proposed a suboptimal transmit solution based on alternating optimization methods. The proposed approach can handle general covariance constraints, including total and per-antenna power constraints.

Besides power constraints, hardware limitations can impose a more stringent constraint, namely, the amplitude constraint. A typical example is optical wireless communication systems in which the data signal is transmitted by the means of modulating the output intensity of light-emitting diodes (LEDs) [16]. Because of the limited dynamic range of typical LEDs, amplitude constraints on the input current signal are necessary to ensure linear electro-optical conversion and avoid nonlinear distortion. In fact, all practical systems transmit codewords and signals that are limited in amplitude. Unfortunately, amplitude constraints are more difficult to handle, and there is no computable channel capacity expression even for the simple scalar channel [17]. Therefore, lower and upper bounds on the channel capacity have been derived [18], [19]. For the multi-user MISO broadcast channel, the authors in [20] studied the transmitter optimization problem based on linear precoding subject to amplitude constraints. For the MISO wiretap channel, secrecy rate maximization via transmit beamforming was considered in [21].

In this paper, we study the design of linear precoders for the two-user MISO BC-CM subject to total and per-antenna power constraints, and also subject to amplitude constraints. After fixing the input distribution, our goal is to find linear precoders that achieve the boundary points of the secrecy rate region. To this end, we formulate the precoder design problem as a weighted secrecy sum rate maximization problem, subject to any of the aforementioned constraints. The resulting problem, however, has a fractional objective function, making it nonconvex and difficult to solve. To circumvent such a difficulty, we first simplify the objective function using a lower bound on the weighted secrecy sum rate. Then, we transform the maximization problem into an equivalent problem with only two variables. We show that the equivalent problem is more tractable and can be solved iteratively with a subgradient search. In each iteration, we solve the dual of a convex *inner* problem to update the value of the *outer* problem, and also to obtain a subgradient vector that specifies the search direction for the next iteration. We characterize a condition under which the obtained solution is guaranteed to be globally optimal. Furthermore, we show that the inner problem can be easily modified to take into account channel uncertainty caused by limited feedback from the receivers. This leads us to the design of *robust linear precoder*s that guarantee a certain *worst-case secrecy rate* performance.

The main contributions of this paper are summarized as follows:

- We propose a practical linear precoding scheme that entails low implementation and computational complexities.
- We transform the precoder design problem, which is a difficult nonconvex problem with $2N$ variables, where N is the number of transmit antennas, into a more tractable problem with only two variables. We show that the resulting problem can be solved iteratively with a subgradient search, where each iteration involves a convex problem that can be efficiently solved.
- Our formulation of the design problem can handle any convex constraints on the channel input, including total and per-antenna power constraints, and amplitude constraints. Such constraints are essential to model hardware limitations in practical systems.
- Furthermore, our formulation of the problem can handle uncertainty in channel information in order to design robust precoders. Unlike conventional encryption techniques, the performance of physical-layer security systems is inherently sensitive to channel estimation errors. Therefore, robust transmission schemes are necessary to alleviate performance sensitivity in practical systems.

To the best of our knowledge, this paper is the first work to consider linear precoding for the two-user MISO BC-CM, subject to per-antenna power or amplitude constraints. Furthermore, it is the first work to consider robust precoding, for the same channel, by taking channel uncertainty into account.

In the remainder of this section, we declare the notation used throughout the paper. The system model, precoding scheme, and transmit constraints are described in detail in Section II. In Section III, we solve the precoder design problem under the premise of perfect channel information. In Section IV, we extend the design problem to its robust counterpart by considering uncertainty in channel information. In Section V, we provide our numerical examples to illustrate the achievable secrecy rate regions of the proposed precoder. We conclude the paper in Section VI.

Notation: The set of N-dimensional real-valued numbers is denoted by $\mathbb{R}^N$, and the set of N-dimensional nonnegative real-valued numbers is denoted by $\mathbb{R}_+^N$. Vector and matrix transposition is denoted by the superscript $^\mathrm{T}$, and the matrix trace is denoted by $\mathrm{Tr}(\cdot)$. The N-dimensional identity matrix is denoted by $\mathbf{I}_N$. We denote the l_2-norm of the vector $\mathbf{x}$ by $\|\mathbf{x}\|_2$, and the Frobenius norm of the matrix $\mathbf{X}$ by $\|\mathbf{X}\|_\mathrm{F}$. We denote the expectation of the random variable X by $\mathbb{E}\{X\}$, the differential entropy of X by $\mathbb{h}(X)$, and the mutual information between X and Y by $\mathbb{I}(X;Y)$. The Gaussian distribution with zero mean and variance σ^2 is denoted by $\mathcal{N}(0,\sigma^2)$, and the uniform distribution over the interval $[-a,a]$ is denoted by $\mathcal{U}[-a,a]$. Finally, we use the subscripts $_1$ and $_2$ to denote relevance to Users 1 and 2, respectively.

II. System Model

In this section, we describe the channel model, the linear precoding scheme, the achievable secrecy rate regions, and the constraints imposed on the transmitted signal vector.

A. The Two-User MISO BC-CM

We study the problem of secret communication between one transmitter and two independent receivers over the Gaussian MISO broadcast channel. The transmitter has $N \geq 2$ antennas[1], and each receiver has a single antenna. In each communication session, the transmitter has two independent confidential messages, one intended for each receiver. The two messages are simultaneously broadcasted, and the transmitter shall ensure that each message can be reliably decoded by its intended receiver, and is kept confidential from the other one.

We assume narrowband transmission over a quasi-static, i.e., nonfading, Gaussian broadcast channel. The transmitted and received baseband signals, as well as the channel gain vectors, are real-valued, i.e., the carrier phase is not modulated. This model is typical for intensity-modulation, direct-detection optical wireless communication systems that utilize LEDs for data transmission[2] (see, for example, the channel model in [21, Section II-B]), and is also applicable to radio frequency (RF) systems utilizing amplitude modulation schemes, such as amplitude-shift keying (where the baseband data symbols are real-valued). Under these assumptions, the signals observed by the two receivers can be expressed as

$$y_1(t) = \mathbf{h}_1^{\mathrm{T}}\mathbf{x}(t) + n_1(t), \tag{1a}$$
$$y_2(t) = \mathbf{h}_2^{\mathrm{T}}\mathbf{x}(t) + n_2(t), \tag{1b}$$

where $\mathbf{x}(t) \in \mathbb{R}^N$ is the transmitted signal vector, $\mathbf{h}_1 \in \mathbb{R}^N$ and $\mathbf{h}_2 \in \mathbb{R}^N$ are the channel gain vectors, and $n_1(t) \sim \mathcal{N}(0, \sigma_1^2)$ and $n_2(t) \sim \mathcal{N}(0, \sigma_2^2)$ are Gaussian noise samples. For notational simplicity, and without loss of generality, we assume that[3] $\sigma_1^2 = \sigma_2^2 = \sigma^2$. We also assume that $\mathbf{h}_1$ and $\mathbf{h}_2$ are linearly independent to ensure that the MISO broadcast channel in (1) is nondegraded.

Let $\boldsymbol{X}$ be an input random vector that satisfies the constraints on the channel input, and Y_1 and Y_2 be the output random variables. Also let $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ be auxiliary random variables. Then, it was shown in [6, Theorem 4] (see also [7, Lemma 2]) that for any joint probability density function (PDF) $p(\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{X}, Y_1, Y_2)$ that can be written as[4]

$$p(\boldsymbol{V}_1, \boldsymbol{V}_2)\, p(\boldsymbol{X}|\boldsymbol{V}_1, \boldsymbol{V}_2)\, p(Y_1, Y_2|\boldsymbol{X}),$$

the rate pair (R_1, R_2) satisfying

$$0 \leq R_1 \leq \mathbb{I}(\boldsymbol{V}_1; Y_1) - \mathbb{I}(\boldsymbol{V}_1; Y_2|\boldsymbol{V}_2) - \mathbb{I}(\boldsymbol{V}_1; \boldsymbol{V}_2), \tag{2a}$$
$$0 \leq R_2 \leq \mathbb{I}(\boldsymbol{V}_2; Y_2) - \mathbb{I}(\boldsymbol{V}_2; Y_1|\boldsymbol{V}_1) - \mathbb{I}(\boldsymbol{V}_1; \boldsymbol{V}_2) \tag{2b}$$

is achievable for the two-user MISO BC-CM in (1). Achievability of the rate pair in (2) was proved based on a double-binning scheme [7, Section IV]. Thus, given a joint PDF $p(\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{X})$, the achievable secrecy rate region can be determined using (2). On the other hand, given a certain constraint on the channel input $\boldsymbol{X}$, it remains unclear how to choose $p(\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{X})$ such that the secrecy rate region is maximized. For the case of total power constraint, it was shown that the secrecy capacity region of the MISO BC-CM in (1) can be characterized in closed-form [7, Theorem 1], and the boundary points are achievable with the S-DPC scheme. This scheme, however, is difficult to implement in practice [9]. In addition, with per-antenna power constraints, there is no closed-form characterization, and apparently the secrecy capacity region can be only found via an exhaustive search over all input covariance matrices that satisfy the per-antenna power constraint. Furthermore, the S-DPC scheme proposed in [7] does not seem to be applicable to the case with amplitude constraints. This motivates us to consider the linear precoding scheme described in the next subsection.

B. Linear Precoding

We study the secrecy performance of the two-user MISO BC-CM in (1) when the transmitted signal vector $\mathbf{x}(t)$ is constructed as

$$\mathbf{x}(t) = \mathbf{w}_1 s_1(t) + \mathbf{w}_2 s_2(t) = \mathbf{W}\mathbf{s}(t), \tag{3}$$

where $\mathbf{w}_1 \in \mathbb{R}^N$ and $\mathbf{w}_2 \in \mathbb{R}^N$ are fixed beamformers, $s_1(t) \in \mathbb{R}$ and $s_2(t) \in \mathbb{R}$ are independent symbols (or codewords) intended for Users 1 and 2, respectively, $\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2]$ is termed as the *precoding matrix*, or simply the *precoder*, and $\mathbf{s}(t) = [s_1(t)\ s_2(t)]^{\mathrm{T}}$ is the vector of transmitted symbols. Although suboptimal, the precoding scheme in (3) is simple to implement. Furthermore, it will enable us to handle per-antenna power or amplitude constraints.

Substituting (3) back into (1), the signals received at both users can be written as

$$y_1(t) = \mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 s_1(t) + \mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2 s_2(t) + n_1(t), \tag{4a}$$
$$y_2(t) = \mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1 s_1(t) + \mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2 s_2(t) + n_2(t). \tag{4b}$$

Let S_1 and S_2 denote the random variable counterparts of $s_1(t)$ and $s_2(t)$, respectively. Then, the transmission scheme in (3) corresponds to choosing

$$\boldsymbol{V}_1 = \mathbf{w}_1 S_1, \quad \boldsymbol{V}_2 = \mathbf{w}_2 S_2, \quad \text{and} \quad \boldsymbol{X} = \boldsymbol{V}_1 + \boldsymbol{V}_2. \tag{5}$$

Substituting from (5) back into (2), the achievable secrecy rate pair in (2) can be written as

$$0 \leq R_1 \leq \mathbb{I}(S_1; Y_1) - \mathbb{I}(S_1; Y_2|S_2), \tag{6a}$$
$$0 \leq R_2 \leq \mathbb{I}(S_2; Y_2) - \mathbb{I}(S_2; Y_1|S_1). \tag{6b}$$

Note that joint encoding is not utilized in (5), i.e., S_1 and S_2 are independent, and thus $\mathbb{I}(\boldsymbol{V}_1; \boldsymbol{V}_2) = \mathbb{I}(S_1; S_2) = 0$.

C. Transmit Constraints and Secrecy Rate Regions

In this subsection, we describe the transmit constraints considered throughout the paper, and derive closed-form expressions for the secrecy rate pair (R_1, R_2).

[1] We use the term *antenna* to refer to general transmit and receive elements. For example, in an optical wireless communication link, the transmit antenna would be an LED, and the receive antenna would be a photodiode.

[2] LEDs are incoherent light sources, i.e., they emit photons with random phases, and thus the carrier phase cannot be modulated.

[3] This assumption can be always satisfied by properly scaling $y_1(t)$ or $y_2(t)$.

[4] In other words, $(\boldsymbol{V}_1, \boldsymbol{V}_2) \to \boldsymbol{X} \to (Y_1, Y_2)$ forms a Markov chain.

1) Total Average Power Constraint: The most common constraint imposed on the input of Gaussian channels is the total average power constraint. Such a constraint is mathematically convenient, and often leads to closed-form solutions. Furthermore, it provides much insight into the performance of the communication system for a given power budget. Mathematically, a total average power constraint $P_{\rm Tot}$ requires the transmitted codewords $\boldsymbol{X}$ to satisfy the inequality

$$\mathrm{Tr}(\mathbb{E}\{\boldsymbol{X}\boldsymbol{X}^{\rm T}\}) \leq P_{\rm Tot}, \tag{7}$$

where $\mathbb{E}\{\boldsymbol{X}\boldsymbol{X}^{\rm T}\}$ is the transmit covariance matrix. An obvious way to comply with the transmission scheme in (3) and satisfy the constraint in (7) is to choose S_1 and S_2 to be independent and identically distributed (i.i.d.) standard Gaussian random variables, that is

$$S_1 \sim \mathcal{N}(0,1), \quad S_2 \sim \mathcal{N}(0,1), \tag{8a}$$

and ensure that the precoder $\mathbf{W}$ satisfies the inequality

$$\|\mathbf{W}\|_{\rm F}^2 \leq P_{\rm Tot}. \tag{8b}$$

Note that our choice of equal variance for the distributions of S_1 and S_2 (both have unity variance) involves no loss of generality because the power allocated to each user can still be adjusted from the entries of the precoding matrix $\mathbf{W}$.

Now, for a given $\mathbf{W}$, and with Gaussian codewords $S_1, S_2 \sim \mathcal{N}(0,1)$, the mutual information terms in (6a) are simply calculated as

$$\mathbb{I}(S_1;Y_1) = \frac{1}{2}\log_2\left(1+\frac{(\mathbf{h}_1^{\rm T}\mathbf{w}_1)^2}{(\mathbf{h}_1^{\rm T}\mathbf{w}_2)^2+\sigma^2}\right), \tag{9a}$$

$$\mathbb{I}(S_1;Y_2|S_2) = \frac{1}{2}\log_2\left(1+\frac{(\mathbf{h}_2^{\rm T}\mathbf{w}_1)^2}{\sigma^2}\right), \tag{9b}$$

where information is measured in (bits/sec/Hz). Similar expressions can be obtained for the corresponding terms in (6b), and thus we end up with the achievable secrecy rate pair

$$R_1 = \frac{1}{2}\left[\log_2\left(1+\frac{(\mathbf{h}_1^{\rm T}\mathbf{w}_1)^2}{(\mathbf{h}_1^{\rm T}\mathbf{w}_2)^2+\sigma^2}\right)\left(\frac{\sigma^2}{(\mathbf{h}_2^{\rm T}\mathbf{w}_1)^2+\sigma^2}\right)\right]^+, \tag{10a}$$

$$R_2 = \frac{1}{2}\left[\log_2\left(1+\frac{(\mathbf{h}_2^{\rm T}\mathbf{w}_2)^2}{(\mathbf{h}_2^{\rm T}\mathbf{w}_1)^2+\sigma^2}\right)\left(\frac{\sigma^2}{(\mathbf{h}_1^{\rm T}\mathbf{w}_2)^2+\sigma^2}\right)\right]^+, \tag{10b}$$

where $[x]^+$ denotes $\max\{x,0\}$.

2) Per-Antenna Average Power Constraint: Despite its simplicity, the total power constraint in (7) is usually not sufficient to capture limitations in practical communication systems. For example, the so-called *digital beamforming*[5] scheme requires a dedicated transmit RF chain for each antenna element[6]. Clearly, each of these chains has its own power budget. Thus, a more realistic approach to model power limitations at the transmitter is to impose an individual power constraint on each RF chain, or, equivalently, on each antenna element, in addition to the total power constraint. A per-antenna average power constraint $P_i, i=1,\ldots,N$, can be expressed as

$$\mathbb{E}\{X_i^2\} \leq P_i, \quad i=1,\ldots,N, \tag{11}$$

where X_i is the ith entry of $\boldsymbol{X}$. Depending on the values of $P_{\rm Tot}$ and $P_1,\ldots,P_N$, one of the constraints in (7) and (11) may become redundant. In particular:

i) If $\sum_{i=1}^N P_i \leq P_{\rm Tot}$, the per-antenna power constraint (11) becomes dominant and (7) can be ignored.
ii) If $P_i \geq P_{\rm Tot}$ for all $i \in \{1,\ldots,N\}$, then (11) is obviously redundant and the total power constraint (7) is sufficient.
iii) If neither of the above two cases holds, both (7) and (11) can be active simultaneously, and thus they should be taken into account.

Similar to the case of dominant total power constraint, we let the codewords S_1 and S_2 be i.i.d. standard Gaussian random variables. Then, in order to satisfy the per-antenna power constraint in (11), the entires of $\mathbf{W}$ should be chosen such that

$$w_{1i}^2 + w_{2i}^2 \leq P_i, \quad i=1,\ldots,N, \tag{12}$$

where w_{1i} and w_{2i} are the ith entries of $\mathbf{w}_1$ and $\mathbf{w}_2$, respectively. Since S_1 and S_2 are Gaussian, the secrecy rate pair expressions in (10) remain valid for any $\mathbf{W}$ satisfying (12).

3) Amplitude Constraints: Amplitude constraints typically arise in the design of intensity modulation systems. In such systems, the data signal is transmitted by the means of modulating the instantaneous output intensity of a noncoherent light source, typically an LED. In order to ensure linear electro-optical conversion, the input current signal in each LED must remain within a certain range, $[-A, A]$, which is specified by the LED characteristics as well as the DC bias applied to the LED [21, Section II-B]. In other words, the input current signal must satisfy the amplitude constraint

$$|X_i| \leq A_i, \quad i=1,\ldots,N. \tag{13}$$

Such a constraint renders the Gaussian distribution infeasible for the channel input. Alternatively, (13) can be fulfilled by choosing the codewords S_1 and S_2 according to the uniform distribution over the interval $[-1,1]$, i.e.,

$$S_1 \sim \mathcal{U}[-1,1], \quad S_2 \sim \mathcal{U}[-1,1], \tag{14a}$$

and choosing the entries of the precoder $\mathbf{W}$ such that they satisfy the constraint

$$|w_{1i}| + |w_{2i}| \leq A_i, \quad i=1,\ldots,N. \tag{14b}$$

The uniform input distribution was used in [18] to obtain a closed-form rate expression for the amplitude-constrained Gaussian channel without secrecy constraints. It was also utilized in [22] to obtain a closed-form secrecy rate expression for the amplitude-constrained Gaussian wiretap channel. Unlike the Gaussian distribution in (8a), the uniform input distribution in (14a), along with Gaussian noise, do not immediately lead to closed-form expressions for $\mathbb{I}(S_1;Y_1)-\mathbb{I}(S_1;Y_2|S_2)$ in (6a), or the similar terms in (6b). Nevertheless, we can lower-bound these terms to obtain closed-form expressions for the secrecy rate pair (R_1, R_2), as follows.

[5] In fact, all the transmission schemes considered in this paper fall into the category of digital (or baseband) beamforming.

[6] Such a constraint is relaxed in the so-called *hybrid beamforming* scheme where the number of RF chains can be smaller than the number of antennas.

First, we rewrite $\mathbb{I}(S_1;Y_1) - \mathbb{I}(S_1;Y_2|S_2)$ as

$$\mathbb{h}(Y_1) - \mathbb{h}(Y_1|S_1) - \mathbb{h}(Y_2|S_2) + \mathbb{h}(Y_2|S_1,S_2). \tag{15}$$

Using the entropy-power inequality [23, Theorem 17.7.3], the differential entropy $\mathbb{h}(Y_1)$ can be lower bounded as

$$\begin{aligned}\mathbb{h}(Y_1) &= \mathbb{h}(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 S_1 + \mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2 S_2 + N_1)\\ &\geq \frac{1}{2}\log_2\left(2^{2\mathbb{h}(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 S_1)} + 2^{2\mathbb{h}(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2 S_2)} + 2^{2\mathbb{h}(N_1)}\right)\\ &= \frac{1}{2}\log_2\left(4(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1)^2 + 4(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + 2\pi e\sigma^2\right). \end{aligned}\tag{16}$$

On the other hand, the conditional differential entropy $\mathbb{h}(Y_1|S_1)$ can be upper bounded by the differential entropy of a Gaussian random variable having equal variance, that is

$$\begin{aligned}\mathbb{h}(Y_1|S_1) &= \mathbb{h}(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2 S_2 + N_1)\\ &\leq \frac{1}{2}\log_2\left(2\pi e\left(\tfrac{1}{3}(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + \sigma^2\right)\right).\end{aligned}\tag{17}$$

Similarly, we have

$$\mathbb{h}(Y_2|S_2) \leq \frac{1}{2}\log_2\left(2\pi e\left(\tfrac{1}{3}(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + \sigma^2\right)\right). \tag{18}$$

We also have

$$\mathbb{h}(Y_2|S_1,S_2) = \mathbb{h}(N_2) = \frac{1}{2}\log_2\left(2\pi e\sigma^2\right). \tag{19}$$

Substituting (16)-(19) back into (15) yields the rate expression

$$R_1 = \left[\frac{1}{2}\log_2\frac{\left(4(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1)^2 + 4(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + 2\pi e\sigma^2\right)\sigma^2}{2\pi e\left(\frac{1}{3}(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2+\sigma^2\right)\left(\frac{1}{3}(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2+\sigma^2\right)}\right]^+. \tag{20a}$$

Similarly, we have

$$R_2 = \left[\frac{1}{2}\log_2\frac{\left(4(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2)^2 + 4(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + 2\pi e\sigma^2\right)\sigma^2}{2\pi e\left(\frac{1}{3}(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2+\sigma^2\right)\left(\frac{1}{3}(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2+\sigma^2\right)}\right]^+. \tag{20b}$$

III. Precoder Design with Perfect Channel Information

In this section, we focus on the design of the precoder $\mathbf{W}$ under the assumption of perfect channel information. We begin with the case of total and per-antenna power constraints. Then, we show in Section III-E that the problem formulation and solution method can be easily modified to handle amplitude constraints.

A. Problem Formulation

By designing $\mathbf{W}$ we mean finding the set of precoding matrices that achieve the boundary of the secrecy rate region characterized by (R_1, R_2). Assuming total and per-antenna power constraints, the design problem can be expressed by the two-objective optimization problem

$$\begin{aligned}\underset{\mathbf{W}}{\text{maximize}}\quad & (R_1,R_2) & (21a)\\ \text{s.t.}\quad & \|\mathbf{W}\|_{\mathrm{F}}^2 \leq P_{\text{Tot}}, & (21b)\\ & w_{1i}^2 + w_{2i}^2 \leq P_i, \quad i=1,\ldots,N, & (21c)\end{aligned}$$

where partial ordering and maximization of the pair (R_1, R_2) are with respect to (w.r.t.) the nonnegative orthant $\mathbb{R}_+^2$ [24, Section 4.7.5]. In the context of multi-objective optimization, a feasible matrix $\mathbf{W}$ that achieves a rate pair on the boundary of the set of all achievable rate pairs is referred to as *Pareto optimal*, and the corresponding secrecy rate pair (R_1, R_2) is a *Pareto optimal pair*. Thus, solving (21) means finding Pareto optimal matrices $\mathbf{W}$.

The standard approach towards solving (21) is to scalarize the objective via a weighted sum [24, Section 4.7.5], that is to replace (R_1, R_2) with $\rho_1 R_1 + \rho_2 R_2$, where the weights $\rho_1 \geq 0$ and $\rho_2 \geq 0$ are free parameters. Different Pareto optimal points can be obtained by adjusting the relative weight ρ_1/ρ_2 to different values between 0 and ∞. This can be carried out by choosing[7] $\rho_1 = \rho$ and $\rho_2 = 1-\rho$, where ρ is a free parameter taking values in the interval $[0,1]$. Thus, for any $\rho \in [0,1]$, we have the weighted secrecy sum rate maximization problem

$$\begin{aligned}\underset{\mathbf{W}}{\text{maximize}}\quad & R_{\text{wsum}}(\rho) & (22a)\\ \text{s.t.}\quad & \|\mathbf{W}\|_{\mathrm{F}}^2 \leq P_{\text{Tot}}, & (22b)\\ & w_{1i}^2 + w_{2i}^2 \leq P_i, \quad i=1,\ldots,N, & (22c)\end{aligned}$$

where

$$R_{\text{wsum}}(\rho) \triangleq \rho R_1 + (1-\rho)R_2 \tag{22d}$$

is the weighted secrecy sum rate. It is clear that solving (22) with $\rho = 1$ corresponds to finding the maximum achievable secrecy rate for User 1 when User 2 is treated as an eavesdropper, while $\rho = 0$ yields the maximum achievable secrecy rate for User 2.

Ideally we would like to solve (22) with the objective $R_{\text{wsum}}(\rho)$ calculated using the rate expressions in (10). Using these expressions, however, will result in a fractional nonconcave objective that is difficult to handle, and the problem in (22) will probably be intractable, except for the special cases $\rho = 0$ and $\rho = 1$. Therefore, we shall simplify the objective of (22) by replacing it with the lower bound

$$\hat{R}_{\text{wsum}}(\rho) = \rho\hat{R}_1 + (1-\rho)\hat{R}_2, \tag{23}$$

where, for $\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 \neq 0$ and $\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2 \neq 0$, $\hat{R}_1$ and $\hat{R}_2$, respectively, are given by

$$\hat{R}_1 = \log_2\frac{\left|\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1\right|\sigma}{((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2+\sigma^2)^{\frac{1}{2}}((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2+\sigma^2)^{\frac{1}{2}}}, \tag{24a}$$

$$\hat{R}_2 = \log_2\frac{\left|\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2\right|\sigma}{((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2+\sigma^2)^{\frac{1}{2}}((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2+\sigma^2)^{\frac{1}{2}}}. \tag{24b}$$

From (10) and (24), it is clear that[8] $\hat{R}_1 < R_1$ and $\hat{R}_2 < R_2$. Thus, for any $\rho \in [0,1]$, we have the inequality $\hat{R}_{\text{wsum}} < R_{\text{wsum}}$. Substituting from (24) into (23), we obtain

$$\hat{R}_{\text{wsum}}(\rho) = \log_2\frac{\left|\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1\right|^{\rho}\left|\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2\right|^{1-\rho}\sigma}{((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2+\sigma^2)^{\frac{1}{2}}((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2+\sigma^2)^{\frac{1}{2}}}. \tag{25}$$

[7] Although constraining ρ_1 and ρ_2 to sum to 1 looks arbitrary here, we will need this assumption in the proof of Proposition 2, particularly to ensure that the objective function in (65) is concave.

[8] The inequality $\hat{R}_1 < R_1$ results from dropping the term 1 in the logarithm in (9a), and dropping the operator $[\cdot]^+$ from the rate expression in (10a). In a similar way, it can be shown that $\hat{R}_2 < R_2$.

Note that $\hat{R}_{\text{wsum}}$ is a tight lower bound for R_{wsum} when the signal-to-noise ratio (SNR) at both receivers is sufficiently high. However, unlike R_{wsum}, whose nonnegativity is ensured by the $[\cdot]^+$ operators in (10), the lower bound $\hat{R}_{\text{wsum}}$ can be negative since $\hat{R}_1$ and/or $\hat{R}_2$ can be negative when the corresponding SNR is sufficiently low. Nonetheless, maximizing $\hat{R}_{\text{wsum}}$ is still beneficial even when its optimal value ends up to be negative because the maximization problem is only used for designing $\mathbf{W}$. The achievable rate pair, however, is obtained by substituting the obtained $\mathbf{W}$ back into (10), i.e., the achievable rate pair is guaranteed to be nonnegative.

Now, we formulate our design problem as[9]

$$\mathbf{W}^\star = \underset{\mathbf{W}}{\text{argmax}} \quad \ln \frac{(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1)^\rho(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2)^{1-\rho}}{((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2+\sigma^2)^{\frac{1}{2}}((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2+\sigma^2)^{\frac{1}{2}}} \tag{26a}$$

$$\text{s.t.} \quad \|\mathbf{W}\|_{\mathrm{F}}^2 \le P_{\text{Tot}}, \tag{26b}$$

$$w_{1i}^2 + w_{2i}^2 \le P_i, \quad i = 1,\ldots,N. \tag{26c}$$

Note that the formulation in (26) implicitly adds the two constraints $\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 \ge 0$ and $\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2 \ge 0$. These additional constraints cause no loss in performance because an optimal $\mathbf{w}_1$ that leads to negative $\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1$ can always be replaced by $-\mathbf{w}_1$ without reducing the optimal value or violating the constraints on $\mathbf{W}$. In a similar way, the implicit constraint $\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2 \ge 0$ can be justified. Note also that, unlike the expressions in (24), the formulation in (26) does not exclude the cases $\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 = 0$ and $\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2 = 0$. For example, the solution $\mathbf{w}_1 = \mathbf{0}$ (which results in $\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 = 0$) would be optimal only when[10] $\rho = 0$, resulting in $(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1)^\rho = 0^0 = 1$.

In the next subsection, we shall explain in detail our approach to solve (26).

B. The Outer Problem

Using the auxiliary variables $\delta_1 \ge 0$ and $\delta_2 \ge 0$, the problem in (26) can be expressed as

$$\underset{\mathbf{W},\delta_1,\delta_2}{\text{maximize}} \quad \ln \frac{(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1)^\rho(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2)^{1-\rho}}{(\delta_1^2+\sigma^2)^{\frac{1}{2}}(\delta_2^2+\sigma^2)^{\frac{1}{2}}} \tag{27a}$$

$$\text{s.t.} \quad |\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1| \le \delta_1, \quad |\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2| \le \delta_2, \tag{27b}$$

$$\|\mathbf{W}\|_{\mathrm{F}}^2 \le P_{\text{Tot}}, \tag{27c}$$

$$w_{1i}^2 + w_{2i}^2 \le P_i, \quad i = 1,\ldots,N. \tag{27d}$$

Let $f(\delta_1,\delta_2)$ denote the optimal value of the *perturbed problem*

$$\underset{\mathbf{W}}{\text{maximize}} \quad \rho\ln(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1) + (1-\rho)\ln(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2) \tag{28a}$$

$$\text{s.t.} \quad |\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1| \le \delta_1, \quad |\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2| \le \delta_2, \tag{28b}$$

$$\|\mathbf{W}\|_{\mathrm{F}}^2 \le P_{\text{Tot}}, \tag{28c}$$

$$w_{1i}^2 + w_{2i}^2 \le P_i, \quad i = 1,\ldots,N. \tag{28d}$$

Then, the problem in (27) can be written as

$$\underset{\delta_1,\delta_2\ge 0}{\text{maximize}} \quad \varphi(\delta_1,\delta_2), \tag{29a}$$

where

$$\varphi(\delta_1,\delta_2) \triangleq f(\delta_1,\delta_2) - \frac{1}{2}\ln\left((\delta_1^2+\sigma^2)(\delta_2^2+\sigma^2)\right). \tag{29b}$$

Now, we can see that solving the design problem in (26) entails solving (28) and (29) iteratively. For obvious reasons, we shall refer to (29) as the *outer problem*, and to (28) as the *inner problem*.

The inner problem is clearly convex, and thus can be efficiently solved using standard convex optimization packages. On the other hand, the outer problem is nonconvex because the objective function $\varphi(\delta_1,\delta_2)$ is not concave, in general. Nevertheless, the following two propositions reveal that $\varphi(\delta_1,\delta_2)$ has a special structure that makes the outer problem solvable, i.e., its global maximum can be efficiently obtained, when a certain condition is satisfied. Even when such a condition is not satisfied, the propositions still give us guidelines for approaching the outer problem.

Proposition 1: The objective function of the outer problem (29) is concave when restricted inside the region

$$\mathcal{D} \triangleq \{(\delta_1,\delta_2) : 0 \le \delta_1 \le \sigma, 0 \le \delta_2 \le \sigma\}.$$

Proof: The proof is fairly straightforward. The first term in (29b), i.e., $f(\delta_1,\delta_2)$, is concave for all $\delta_1,\delta_2 \ge 0$ because the perturbed problem (28) is convex [24, Section 5.6.1]. On the other hand, the second term $-\frac{1}{2}\ln\left((\delta_1^2+\sigma^2)(\delta_2^2+\sigma^2)\right)$ is concave only when $0 \le \delta_1 \le \sigma$ and $0 \le \delta_2 \le \sigma$ (this can be easily verified after writing down the Hessian matrix). Thus, $\varphi(\delta_1,\delta_2)$ is concave when $(\delta_1,\delta_2) \in \mathcal{D}$. ■

Proposition 2: The objective function of the outer problem (29) is quasiconcave when restricted to any line (in the nonnegative orthant $\mathbb{R}_+^2$) passing through the origin.

The proof, which is provided in Appendix A, is based on the observation that the first term in (29b) is nondecreasing w.r.t. $\mathbb{R}_+^2$, while the second term is monotonically decreasing[11]. Note that the condition in Proposition 2 is weaker than stating that $\varphi(\delta_1,\delta_2)$ is quasiconcave, as the latter condition would require φ to be quasiconcave when restricted to *any line* in $\mathbb{R}_+^2$.

Now, if $\varphi(\delta_1,\delta_2)$ has a maximum inside $\mathcal{D}$, then it is the only maximum inside $\mathcal{D}$ according to Proposition 1. This also implies that $\varphi(\delta_1,\delta_2)$ has one maximum inside $\mathcal{D}$ when restricted to any line (inside $\mathcal{D}$) passing through the origin. However, as per Proposition 2, $\varphi(\delta_1,\delta_2)$ can only have one maximum (inside and outside $\mathcal{D}$) when it is restricted to any line passing through the origin. Thus, combining Propositions 1 and 2 yields the following conclusion:

Corollary 1: For the outer problem (29), any local maximum inside the region $\mathcal{D}$ is a global maximum.

Corollary 1 suggests that we begin searching for the solution of (29) inside $\mathcal{D}$. If the search algorithm terminates at $\boldsymbol{\delta}^\star \in \mathcal{D}$, then $\boldsymbol{\delta}^\star$ is guaranteed to be the (globally) optimal solution of (29). On the other hand, if $\boldsymbol{\delta}^\star \notin \mathcal{D}$, then we will accept

[9] Using the natural logarithm $\ln(\cdot)$ in the objective of (26) instead of $\log_2(\cdot)$ will simplify the notation when differentiation becomes involved.

[10] This is true because we assume that $\mathbf{h}_1$ and $\mathbf{h}_2$ are linearly independent. On the other hand, if $\mathbf{h}_1$ and $\mathbf{h}_2$ are collinear and $\|\mathbf{h}_1\|_2 \le \|\mathbf{h}_2\|_2$, then $\mathbf{w}_1 = \mathbf{0}$ would be optimal for all $\rho \in [0,1]$, i.e., User 1 cannot achieve positive secrecy rates and should always be treated as an eavesdropper because its channel $\mathbf{h}_1$ is degraded.

[11] See [24, Section 3.6.1] for the notion of *monotonicity w.r.t. a generalized inequality* on the nonnegative orthant.

$\boldsymbol{\delta}^\star$ as a (possibly) suboptimal solution. It is worth to mention that the numerical results show that $\varphi(\delta_1, \delta_2)$ is a *unimodal* function with only one maximum, for all $\delta_1, \delta_2 \geq 0$, and no other stationary points. However, it is difficult, in general, to rigorously prove that a multivariable function is unimodal, beyond concavity or quasiconcavity. Therefore, we can only conjecture that $\varphi(\delta_1, \delta_2)$ is unimodal (for all $\delta_1, \delta_2 \geq 0$), and consequently any local maximum is global.

Now, we have to choose a reasonable search algorithm to solve (29). Since the objective function $\varphi(\delta_1, \delta_2)$ is differentiable almost everywhere (because $f(\delta_1, \delta_2)$ is differentiable almost everywhere), a natural choice for the search algorithm is the *subgradient method* in which the subgradient vectors are used as the search directions [25], [26]. Let the vector $\nabla_{\text{sub}} f(\delta_1, \delta_2) \in \mathbb{R}^2_+$ be a subgradient[12] of f at (δ_1, δ_2), where the two entries of $\nabla_{\text{sub}} f$ are both nonnegative since f is nondecreasing w.r.t. δ_1 and δ_2. Then, from (29b), the corresponding subgradient of φ is given by

$$\nabla_{\text{sub}}\varphi(\delta_1, \delta_2) = \nabla_{\text{sub}} f(\delta_1, \delta_2) - \begin{bmatrix} \dfrac{\delta_1}{\delta_1^2 + \sigma^2} & \dfrac{\delta_2}{\delta_2^2 + \sigma^2} \end{bmatrix}^{\mathrm{T}}. \tag{30}$$

Before we proceed to the details of the search algorithm, we need to find $\nabla_{\text{sub}} f$ in order to calculate the search direction $\nabla_{\text{sub}}\varphi$ at any (δ_1, δ_2). This will be our goal in the next subsection.

C. The Dual of the Inner Problem

The inner problem (28) is a convex problem whose constraints satisfy Slater's condition, and thus strong duality holds [24, Section 5.2.3]. As a consequence, the optimal value of the inner problem, i.e., $f(\delta_1, \delta_2)$, is identical to the optimal value of its (Lagrange) dual. Furthermore, the optimal Lagrange multipliers associated with the two constraints in (28b) provide a subgradient vector for f at (δ_1, δ_2) [27, Section 8.5.6]. Therefore, our next task is to derive the dual problem for (28).

We begin with reformulating (28) as

$$\begin{aligned}
\underset{\mathbf{W}, z_1, \ldots, z_4}{\text{maximize}} \quad & \rho \ln z_1 + (1-\rho) \ln z_2 && \text{(31a)}\\
\text{s.t.} \quad & |z_3| \leq \delta_1, \quad |z_4| \leq \delta_2, && \text{(31b)}\\
& \|\mathbf{w}_1\|_2^2 + \|\mathbf{w}_2\|_2^2 \leq P_{\text{Tot}}, && \text{(31c)}\\
& w_{1i}^2 + w_{2i}^2 \leq P_i, \quad i = 1, \ldots, N, && \text{(31d)}\\
& \mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 = z_1, \quad \mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2 = z_2, && \text{(31e)}\\
& \mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1 = z_3, \quad \mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2 = z_4, && \text{(31f)}
\end{aligned}$$

where we have introduced four new variables, $z_1, \ldots, z_4$, and four associated equality constraints in (31e)-(31f). The Lagrangian associated with the reformulated problem (31) is

$$\begin{aligned}
&L(\mathbf{W}, z_1, \ldots, z_4, \lambda_1, \lambda_2, \gamma, \boldsymbol{\mu}, \nu_1, \ldots, \nu_4)\\
&\quad = \rho \ln z_1 + (1-\rho)\ln z_2\\
&\qquad - \lambda_1 (|z_3| - \delta_1) - \lambda_2 (|z_4| - \delta_2)\\
&\qquad - \gamma \left(\|\mathbf{w}_1\|_2^2 + \|\mathbf{w}_2\|_2^2 - P_{\text{Tot}}\right)\\
&\qquad - \sum_{i=1}^{N} \mu_i (w_{1i}^2 + w_{2i}^2 - P_i)\\
&\qquad - \nu_1(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 - z_1) - \nu_2(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2 - z_2),\\
&\qquad - \nu_3(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1 - z_3) - \nu_4(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2 - z_4),
\end{aligned} \tag{32}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are the Lagrange multipliers associated with the perturbed constraints in (31b), $\gamma \geq 0$ is the Lagrange multiplier associated with the total power constraint (31c), $\boldsymbol{\mu} = [\mu_1 \ldots \mu_N]^{\mathrm{T}}$, with entries $\mu_i \geq 0, i = 1, \ldots, N$, is the Lagrange multiplier vector associated with the per-antenna power constraint (31d), and $\nu_1, \ldots, \nu_4$ are the Lagrange multipliers associated with the equality constraints in (31e)-(31f). Upon rearranging the terms in the Lagrangian (32), the dual function g is obtained by maximization over the primary variables $\mathbf{W}, z_1, \ldots, z_4$, that is

$$\begin{aligned}
&g(\lambda_1, \lambda_2, \gamma, \boldsymbol{\mu}, \nu_1, \ldots, \nu_4)\\
&\quad = \lambda_1\delta_1 + \lambda_2\delta_2 + \gamma P_{\text{Tot}} + \sum_{i=1}^{N} \mu_i P_i\\
&\qquad + \sum_{i=1}^{N} \max_{w_{1i}} \left(-\left(\nu_1 h_{1i} + \nu_3 h_{2i}\right) w_{1i} - (\gamma + \mu_i) w_{1i}^2\right)\\
&\qquad + \sum_{i=1}^{N} \max_{w_{2i}} \left(-\left(\nu_2 h_{2i} + \nu_4 h_{1i}\right) w_{2i} - (\gamma + \mu_i) w_{2i}^2\right)\\
&\qquad + \max_{z_1} \left(\nu_1 z_1 + \rho \ln z_1\right)\\
&\qquad + \max_{z_2} \left(\nu_2 z_2 + (1-\rho)\ln z_2\right)\\
&\qquad + \max_{z_3} \left(\nu_3 z_3 - \lambda_1 |z_3|\right) + \max_{z_4} \left(\nu_4 z_4 - \lambda_2 |z_4|\right),
\end{aligned} \tag{33}$$

where h_{1i} and h_{2i} are the ith entries of $\mathbf{h}_1$ and $\mathbf{h}_2$, respectively. Now, we have to solve all the maximization terms in (33) analytically. In fact, we have

$$\begin{aligned}
&\max_{w_{1i}} \left(-\left(\nu_1 h_{1i} + \nu_3 h_{2i}\right) w_{1i} - (\gamma + \mu_i) w_{1i}^2\right)\\
&\qquad = \frac{\left(\nu_1 h_{1i} + \nu_3 h_{2i}\right)^2}{4(\gamma + \mu_i)}, \quad \gamma + \mu_i > 0, \quad i = 1, \ldots, N,
\end{aligned} \tag{34a}$$

$$\max_{z_1} \left(\nu_1 z_1 + \rho \ln z_1\right) = -\rho \ln \frac{-\nu_1}{\rho} - \rho, \quad \nu_1 < 0, \tag{34b}$$

$$\max_{z_3} \left(\nu_3 z_3 - \lambda_1 |z_3|\right) = \begin{cases} 0 & |\nu_3| \leq \lambda_1 \\ \infty & \text{otherwise} \end{cases}, \tag{34c}$$

where (34a) is a simple unconstrained quadratic concave maximization problem, (34b) follows from the conjugate of the negative logarithm function (see [24, Example 3.21]), and (34c) follows from the conjugate of the absolute value function (see [24, Example 3.26]). Note that the condition $\gamma + \mu_i > 0$ in (34a) is always satisfied because, for each antenna, at least one of the constraints (i.e., the total power

[12] Perhaps the term *supergradient* would be more appropriate here because f is a concave function.

constraint or the per-antenna power constraint) must be active. Thus, $\gamma + \mu_i$ is strictly positive for all $i = 1, \ldots, N$. Using the expressions in (34), the dual problem can be formulated as[13]

$$\underset{\substack{\lambda_1,\lambda_2,\gamma,\boldsymbol{\mu},\boldsymbol{\tau}_1,\\ \boldsymbol{\tau}_2,\nu_1,\ldots,\nu_4}}{\text{minimize}} \begin{pmatrix} \delta_1\lambda_1 + \delta_2\lambda_2 + P_{\text{Tot}}\gamma \\ +\sum_{i=1}^{N}(P_i\mu_i + \tau_{1i} + \tau_{2i}) \\ -\rho \ln \dfrac{-\nu_1}{\rho} - (1-\rho)\ln\dfrac{-\nu_2}{1-\rho} \end{pmatrix} - 1 \quad (35a)$$

$$\text{s.t.} \quad \nu_1, \nu_2 < 0, \quad |\nu_3| \le \lambda_1, \quad |\nu_4| \le \lambda_2, \quad (35b)$$

$$\gamma \ge 0, \quad \mu_i \ge 0, \quad \gamma + \mu_i > 0, \quad (35c)$$

$$\begin{bmatrix} \tau_{1i} & \nu_1 h_{1i} + \nu_3 h_{2i} \\ \nu_1 h_{1i} + \nu_3 h_{2i} & 4(\gamma + \mu_i) \end{bmatrix} \succeq 0, \quad (35d)$$

$$\begin{bmatrix} \tau_{2i} & \nu_2 h_{2i} + \nu_4 h_{1i} \\ \nu_2 h_{2i} + \nu_4 h_{1i} & 4(\gamma + \mu_i) \end{bmatrix} \succeq 0, \quad (35e)$$

$$i = 1, \ldots, N,$$

where we have used *Schur complement*, in conjunction with the auxiliary variables τ_{1i} and τ_{2i}, $i = 1, \ldots, N$, to formulate the linear matrix inequality constraints in (35d) and (35e). Two special cases of the dual problem (35) are worth mentioning.

Firstly, at the corner point $\rho = 0$, the Lagrange multipliers λ_1, ν_1, and ν_3 are set to zero, and the dual problem (35) simplifies to

$$\underset{\substack{\lambda_2,\gamma,\boldsymbol{\mu},\\ \boldsymbol{\tau}_2,\nu_2,\nu_4}}{\text{minimize}} \begin{pmatrix} \delta_2\lambda_2 + P_{\text{Tot}}\gamma \\ +\sum_{i=1}^{N}(P_i\mu_i + \tau_{2i}) - \ln(-\nu_2) \end{pmatrix} - 1 \quad (36a)$$

$$\text{s.t.} \quad \nu_2 < 0, \quad |\nu_4| \le \lambda_2, \quad (36b)$$

$$\gamma \ge 0, \quad \mu_i \ge 0, \quad \gamma + \mu_i > 0, \quad (36c)$$

$$\begin{bmatrix} \tau_{2i} & \nu_2 h_{2i} + \nu_4 h_{1i} \\ \nu_2 h_{2i} + \nu_4 h_{1i} & 4(\gamma + \mu_i) \end{bmatrix} \succeq 0, \quad (36d)$$

$$i = 1, \ldots, N,$$

where we have used the convention that $0 \ln \frac{0}{0} = 0$ while simplifying the objective function. Similar simplification can be obtained for the other corner point (i.e., at $\rho = 1$).

Secondly, for the case in which there is only a total power constraint, i.e., when (31d) does not exist or is not active, the Lagrange multiplier vector $\boldsymbol{\mu}$ is set to $\mathbf{0}$, and (35) simplifies to

$$\underset{\substack{\lambda_1,\lambda_2,\gamma,\tau_1,\\ \tau_2,\nu_1,\ldots,\nu_4}}{\text{minimize}} \begin{pmatrix} \delta_1\lambda_1 + \delta_2\lambda_2 + P_{\text{Tot}}\gamma + \tau_1 + \tau_2 \\ -\rho \ln \dfrac{-\nu_1}{\rho} - (1-\rho)\ln\dfrac{-\nu_2}{1-\rho} \end{pmatrix} - 1 \quad (37a)$$

$$\text{s.t.} \quad \nu_1, \nu_2 < 0, \quad |\nu_3| \le \lambda_1, \quad |\nu_4| \le \lambda_2, \quad (37b)$$

$$\gamma > 0, \quad (37c)$$

$$\begin{bmatrix} \tau_1 & (\nu_1 \mathbf{h}_1 + \nu_3 \mathbf{h}_2)^{\mathrm{T}} \\ \nu_1 \mathbf{h}_1 + \nu_3 \mathbf{h}_2 & 4\gamma \mathbf{I}_N \end{bmatrix} \succeq 0, \quad (37d)$$

$$\begin{bmatrix} \tau_2 & (\nu_2 \mathbf{h}_2 + \nu_4 \mathbf{h}_1)^{\mathrm{T}} \\ \nu_2 \mathbf{h}_2 + \nu_4 \mathbf{h}_1 & 4\gamma \mathbf{I}_N \end{bmatrix} \succeq 0. \quad (37e)$$

The dual problem (35) is a semidefinite program [28]. Thus, it can be efficiently solved using the interior-point method with a worst-case complexity of [29]

$$\mathcal{O}\left(\max\{n, m\}^4 n^{\frac{1}{2}} \log(1/\epsilon)\right),$$

where n is the number of variables, m is the number of constraints, and ϵ is the desired accuracy of the solution. In practice, semidefinite programs are conveniently and efficiently solved using standard convex optimization packages, such as CVX [30] and MOSEK [31]. Therefore, (35) can be efficiently solved to obtain $f(\delta_1, \delta_2)$ and $\nabla_{\text{sub}} f(\delta_1, \delta_2)$. Let $\{\lambda_1^\star, \lambda_2^\star, \gamma^\star, \boldsymbol{\mu}^\star, \boldsymbol{\tau}_1^\star, \boldsymbol{\tau}_2^\star, \nu_1^\star, \ldots, \nu_4^\star\}$ denote the optimal solution of (35) for fixed δ_1 and δ_2. Then, $f(\delta_1, \delta_2)$ is equal to the optimal value of the objective, and the vector $[\lambda_1^\star \ \lambda_2^\star]^{\mathrm{T}}$ is a subgradient of f at (δ_1, δ_2). Consequently, the subgradient vector in (30) can be written as

$$\nabla_{\text{sub}} \varphi(\delta_1, \delta_2) = \left[\lambda_1^\star - \frac{\delta_1}{\delta_1^2 + \sigma^2} \quad \lambda_2^\star - \frac{\delta_2}{\delta_2^2 + \sigma^2}\right]^{\mathrm{T}}. \quad (38)$$

Having obtained $\nabla_{\text{sub}} \varphi(\delta_1, \delta_2)$, we are now ready to solve the outer problem (29).

D. The Search Algorithm

In this subsection, we turn our focus to the search algorithm used to find a solution for the outer problem (29), that is to find $\boldsymbol{\delta}^\star = [\delta_1^\star \ \delta_2^\star]^{\mathrm{T}}$ that maximizes $\varphi(\delta_1, \delta_2)$. A typical subgradient method uses the iteration [26]

$$\boldsymbol{\delta}^{(k+1)} = \boldsymbol{\delta}^{(k)} + \alpha^{(k)} \, \nabla_{\text{sub}} \varphi(\boldsymbol{\delta}^{(k)}), \quad k = 1, 2, \ldots, \quad (39)$$

where $\boldsymbol{\delta}^{(k)}$ is the start point at the kth iteration (with $\boldsymbol{\delta}^{(1)}$ being the initial point), $\alpha^{(k)} > 0$ is the kth step size, and $\boldsymbol{\delta}^{(k+1)}$ is the end point after k iterations. The numerical results in Section V reveal that, when the noise variance σ^2 is equal to 1, the values of $\delta_1^\star$ and $\delta_2^\star$ can be on the order of 10^{-3} up to 10^1. This several orders of magnitude difference suggests that the search is better carried out on a logarithmic scale, rather than the ordinary linear scale, in order to improve the accuracy and maintain numerical stability (so convergence is achieved within a reasonable number of iterations).

Let $\boldsymbol{\delta}_{\text{dB}}$ be defined as $\boldsymbol{\delta}_{\text{dB}} \triangleq [20\log_{10}(\delta_1) \quad 20\log_{10}(\delta_2)]^{\mathrm{T}}$. Then, the subgradient $\nabla_{\text{sub}}\varphi$ on the logarithmic scale, i.e., when differentiation is w.r.t. $20\log_{10}(\delta_1)$ and $20\log_{10}(\delta_2)$, is given by

$$\nabla_{\text{sub}} \varphi(\boldsymbol{\delta}_{\text{dB}}) = \frac{\ln 10}{20} \begin{bmatrix} \delta_1 \left(\lambda_1^\star - \dfrac{\delta_1}{\delta_1^2 + \sigma^2}\right) \\ \delta_2 \left(\lambda_2^\star - \dfrac{\delta_2}{\delta_2^2 + \sigma^2}\right) \end{bmatrix}. \quad (40)$$

Now, we proceed with the search algorithm as follows. First, we choose an initial point $\boldsymbol{\delta}_{\text{dB}}^{(1)}$, such that $\delta_1^{(1)} \le \sigma$ and $\delta_2^{(1)} \le \sigma$. This point is iteratively updated by

$$\boldsymbol{\delta}_{\text{dB}}^{(k+1)} = \boldsymbol{\delta}_{\text{dB}}^{(k)} + \alpha_{\text{dB}}^{\text{Fix}} \frac{\nabla_{\text{sub}} \varphi(\boldsymbol{\delta}_{\text{dB}}^{(k)})}{\|\nabla_{\text{sub}} \varphi(\boldsymbol{\delta}_{\text{dB}}^{(k)})\|_2}, \quad k = 1, 2, \ldots, \quad (41)$$

where $\alpha_{\text{dB}}^{\text{Fix}}$ is a fixed step size in dB. That is, for each iteration, we take a step $\alpha_{\text{dB}}^{\text{Fix}}$ in the direction of the subgradient. This

[13] We maintain the fixed term -1 in the objective function in (35a) to have its optimal value equal to the optimal value of the inner problem (28), i.e., equal to $f(\delta_1, \delta_2)$.

TABLE I
SUBGRADIENT-BASED SEARCH ALGORITHM TO SOLVE (29).

Algorithm 1 A subgradient-based algorithm to solve (29)

```
Set the initial (fixed) step size $\alpha_{\mathrm{dB}}^{\mathrm{Fix}}$ and the maximum number of iterations with decreasing step size $L$
Set the binary switch REDUCE = **false**    ▷ "REDUCE" is a switch that determines whether to proceed with a "fixed" or "decreasing" step
Set the indexes $k = 1$ and $l = 0$
Choose an initial point $\boldsymbol{\delta}_{\mathrm{dB}}^{(1)}$ such that $\delta_1^{(1)} \le \sigma$, $\delta_2^{(1)} \le \sigma$
while $l \le L$ do
    Solve (35) to obtain $f(\boldsymbol{\delta}_{\mathrm{dB}}^{(k+l)})$, $\lambda_1^{\star(k+l)}$, $\lambda_2^{\star(k+l)}$
    Calculate $\varphi(\boldsymbol{\delta}_{\mathrm{dB}}^{(k+l)})$ using (29b)
    Calculate $\nabla_{\mathrm{sub}}\varphi(\boldsymbol{\delta}_{\mathrm{dB}}^{(k+l)})$ using (40)
    if $k \ge 2$
        if $\varphi(\boldsymbol{\delta}_{\mathrm{dB}}^{(k+l)}) \le \varphi(\boldsymbol{\delta}_{\mathrm{dB}}^{(k+l-1)})$, then
            REDUCE = true
        end if
    end if
    if REDUCE = false, then
        Update $\boldsymbol{\delta}_{\mathrm{dB}}^{(k+l)}$ using (41)
        $k := k + 1$
    else
        Update $\boldsymbol{\delta}_{\mathrm{dB}}^{(k+l)}$ using (42)
        $l := l + 1$
    end if
end while
return $\boldsymbol{\delta}_{\mathrm{dB}}^{\star} = \mathrm{argmax}\,\{\varphi(\boldsymbol{\delta}_{\mathrm{dB}}^{(1)}), \dots, \varphi(\boldsymbol{\delta}_{\mathrm{dB}}^{(k+L)})\}$
```

iteration shall continue until we overshoot the peak, i.e., when $\varphi(\boldsymbol{\delta})$ starts to decrease. Once the peak is encountered, we reduce the step size and use the new iteration

$$\boldsymbol{\delta}_{\mathrm{dB}}^{(K+l+1)} = \boldsymbol{\delta}_{\mathrm{dB}}^{(K+l)} + \frac{\alpha_{\mathrm{dB}}^{\mathrm{Fix}}}{l}\,\frac{\nabla_{\mathrm{sub}}\varphi(\boldsymbol{\delta}_{\mathrm{dB}}^{(K+l)})}{\|\nabla_{\mathrm{sub}}\varphi(\boldsymbol{\delta}_{\mathrm{dB}}^{(K+l)})\|_2},\ l = 1, \dots, L, \tag{42}$$

where K is the number of iterations using (41), i.e., with a fixed step size, and L is the maximum number of iterations with a decreasing step size. Unlike K, L shall be determined in advance according to the required accuracy of the solution. Therefore, the search will terminate after $K+L$ total iterations, and the solution $\boldsymbol{\delta}_{\mathrm{dB}}^{\star}$ is obtained with accuracy $\alpha_{\mathrm{dB}}^{\mathrm{Fix}}/L$ dB. For convenience, the search algorithm is summarized in Table I.

Upon solving the outer problem (29), we solve the inner problem (28) using $\boldsymbol{\delta}^{\star}$ to obtain the optimum precoder $\mathbf{W}^{\star}$. Then, the secrecy rate pair (R_1, R_2) is calculated by substituting $\mathbf{W}^{\star}$ into (10). We repeat the entire procedure with different values of $\rho \in [0, 1]$ to obtain different points (R_1, R_2) on the boundary of the achievable secrecy rate region. The numerical results in Section V show that the outer problem is solved with Algorithm 1 in about 20–30 iterations. In each iteration, the main computational cost comes from solving the dual problem (35) to obtain a subgradient vector. Thus, the overall computational complexity is determined by the complexity of (35), which is a convex semidefinite problem, times the number of iterations required for the outer problem.

E. Per-Antenna Amplitude Constraint

In this subsection, we design the precoding matrix $\mathbf{W}$ subject to the per-antenna amplitude constraint (14b). Fortunately, the problem formulation and solution technique developed in the previous subsections are immediately applicable. In fact, we just need to modify the weighted secrecy sum rate expression (25) and the inner problem (28), and consequently its dual (35), to take the amplitude constraint into account.

Similar to (25), we need a weighted secrecy sum rate expression that is amenable to optimization. From (20), the rate expressions R_1 and R_2, respectively, can be lower-bounded by

$$\hat{R}_1 = \log_2 \frac{3\sqrt{2}\,\left|\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1\right|\sigma}{\sqrt{\pi e}\,\left((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + 3\sigma^2\right)^{\frac{1}{2}}\left((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + 3\sigma^2\right)^{\frac{1}{2}}}, \tag{43a}$$

$$\hat{R}_2 = \log_2 \frac{3\sqrt{2}\,\left|\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2\right|\sigma}{\sqrt{\pi e}\,\left((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + 3\sigma^2\right)^{\frac{1}{2}}\left((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + 3\sigma^2\right)^{\frac{1}{2}}}. \tag{43b}$$

Then, for any $\rho \in [0, 1]$, we have the weighted secrecy sum rate

$$\hat{R}_{\mathrm{wsum}}(\rho) = \log_2 \frac{3\sqrt{2}\,\left|\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1\right|^{\rho}\left|\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2\right|^{1-\rho}\sigma}{\sqrt{\pi e}((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + 3\sigma^2)^{\frac{1}{2}}((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + 3\sigma^2)^{\frac{1}{2}}}. \tag{44}$$

Similar to (28), we formulate the inner problem as

$$\underset{\mathbf{W}}{\text{maximize}}\quad \rho\ln(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1) + (1-\rho)\ln(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2) \tag{45a}$$

$$\text{s.t.}\quad |\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1| \le \delta_1, \quad |\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2| \le \delta_2, \tag{45b}$$

$$|w_{1i}| + |w_{2i}| \le A_i, \quad i = 1, \dots, N. \tag{45c}$$

Then, following the same procedure as in Section III-C, it can be shown that the dual problem for (45) is

$$\underset{\substack{\lambda_1,\lambda_2,\boldsymbol{\mu},\\ \nu_1,\dots,\nu_4}}{\text{minimize}} \left(\begin{array}{c} \delta_1\lambda_1 + \delta_2\lambda_2 + \sum_{i=1}^{N}(A_i\mu_i) \\ -\rho\ln\dfrac{-\nu_1}{\rho} - (1-\rho)\ln\dfrac{-\nu_2}{1-\rho} \end{array}\right) - 1 \tag{46a}$$

$$\text{s.t.}\quad \nu_1, \nu_2 < 0, \quad |\nu_3| \le \lambda_1, \quad |\nu_4| \le \lambda_2, \tag{46b}$$

$$|\nu_1 h_{1i} + \nu_3 h_{2i}| \le \mu_i, \quad i = 1, \dots, N, \tag{46c}$$

$$|\nu_2 h_{2i} + \nu_4 h_{1i}| \le \mu_i, \quad i = 1, \dots, N, \tag{46d}$$

where the Lagrange multipliers $\lambda_1, \lambda_2, \nu_1, \dots, \nu_4$ are defined as in (35), and $\boldsymbol{\mu} = [\mu_1 \dots \mu_N]^{\mathrm{T}}$ is the Lagrange multiplier vector associated with the amplitude constraint (45c). The dual problem (46) is, of course, convex. The objective (46a) is a sum of linear functions and negative logarithms, and the constraints involve absolute value functions. Thus, the dual problem (46) can be conveniently solved using CVX [30].

IV. Robust Precoder Design with Imperfect Channel Information

Our solutions in Section III were based on the assumption that the channel gain vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ are precisely known to the transmitter. In this section, we capitalize on our approach and tackle the more general design problem in which the transmitter has only uncertain estimates of $\mathbf{h}_1$ and $\mathbf{h}_2$. We will see that the problem formulation is very similar to its non-robust counterpart, and thus the solution approach will also be similar. Therefore, our pace in this section will be relatively fast, and much of the details and derivations encountered in the previous section will be omitted for brevity.

A. Channel Uncertainty Model

We adopt the spherical uncertainty model (or norm-bounded error model) in which the actual channel gain vectors, $\mathbf{h}_1$ and $\mathbf{h}_2$, respectively, are modeled by

$$\mathbf{h}_1 \in \mathcal{H}_1, \quad \mathcal{H}_1 = \left\{ \hat{\mathbf{h}}_1 + \mathbf{e}_1 : \|\mathbf{e}_1\|_2 \le \epsilon_1 \right\}, \tag{47a}$$

$$\mathbf{h}_2 \in \mathcal{H}_2, \quad \mathcal{H}_2 = \left\{ \hat{\mathbf{h}}_2 + \mathbf{e}_2 : \|\mathbf{e}_2\|_2 \le \epsilon_2 \right\}, \tag{47b}$$

where $\mathcal{H}_1$ and $\mathcal{H}_2$ are N-dimensional spherical sets, $\hat{\mathbf{h}}_1 \in \mathbb{R}^N$ and $\hat{\mathbf{h}}_2 \in \mathbb{R}^N$ are the channel vector estimates available to the transmitter, $\mathbf{e}_1 \in \mathbb{R}^N$ and $\mathbf{e}_2 \in \mathbb{R}^N$ are unknown (but norm-bounded) error vectors, and ϵ_1 and ϵ_2 are known constants that quantify the amount of uncertainty for each channel. This error model is well accepted for representing channel uncertainty caused by quantization errors and finite-rate feedback from the receiver to the transmitter [32, Lemma 1].

Given the uncertain channel information in (47), our goal in this section is to design the precoder $\mathbf{W}$ in order to optimize the performance in terms of the worst-case secrecy rate pair $(R_1^{\text{wc}}, R_2^{\text{wc}})$, that is to solve the two-objective optimization problem

$$\underset{\mathbf{W}}{\text{maximize}} \quad (R_1^{\text{wc}}, R_2^{\text{wc}}) \tag{48}$$

subject to power or amplitude constraints, where, for any $\mathbf{W}$, the worst-case secrecy rates R_1^{wc} and R_2^{wc} are determined by

$$R_1^{\text{wc}} = \min_{\substack{\mathbf{h}_1 \in \mathcal{H}_1,\\ \mathbf{h}_2 \in \mathcal{H}_2}} R_1, \tag{49a}$$

$$R_2^{\text{wc}} = \min_{\substack{\mathbf{h}_1 \in \mathcal{H}_1,\\ \mathbf{h}_2 \in \mathcal{H}_2}} R_2. \tag{49b}$$

Similar to our approach in the previous section, we shall tackle (48) by solving a weighted worst-case secrecy sum rate maximization problem, as we see in the following two subsections.

B. Robust Precoder Design Subject to Total and Per-Antenna Average Power Constraints

In this subsection, we solve the weighted worst-case sum rate maximization problem subject to total and per-antenna power constraints. First, we need to simplify the worst-case secrecy rate expressions in order to obtain a weighted sum rate that is amenable to optimization. Substituting from (10a) into (49a), we obtain

$$\begin{aligned}
R_1^{\text{wc}} &= \Bigg[\frac{1}{2}\log_2 \min_{\mathbf{h}_1 \in \mathcal{H}_1} \left(1 + \frac{(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1)^2}{(\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + \sigma^2} \right) \\
&\qquad + \frac{1}{2}\log_2 \min_{\mathbf{h}_2 \in \mathcal{H}_2} \left(\frac{\sigma^2}{(\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + \sigma^2} \right) \Bigg]^+ \\
&\ge \frac{1}{2}\log_2 \left(1 + \frac{\min\limits_{\mathbf{h}_1 \in \mathcal{H}_1} (\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1)^2}{\max\limits_{\mathbf{h}_1 \in \mathcal{H}_1} (\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + \sigma^2} \right) \\
&\qquad + \frac{1}{2}\log_2 \left(\frac{\sigma^2}{\max\limits_{\mathbf{h}_2 \in \mathcal{H}_2} (\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + \sigma^2} \right) \qquad (50a) \\
&\ge \log_2 \frac{\min\limits_{\mathbf{h}_1 \in \mathcal{H}_1} \left|\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1\right| \sigma}{\max\limits_{\substack{\mathbf{h}_1 \in \mathcal{H}_1,\\ \mathbf{h}_2 \in \mathcal{H}_2}} \left((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + \sigma^2\right)^{\frac{1}{2}} \left((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + \sigma^2\right)^{\frac{1}{2}}}, \qquad (50b)
\end{aligned}$$

where the first inequality follows from dropping the $[\cdot]^+$ operator and applying the inequality

$$\min_x \frac{f_1(x)}{f_2(x)} \ge \frac{\min\limits_x f_1(x)}{\max\limits_x f_2(x)},$$

which holds for any f_1 and f_2, and the second inequality follows from dropping the term 1. We shall use (50b) to formulate the weighted secrecy sum rate for the optimization problem, while we use the better bound in (50a) to calculate the worst-case secrecy rate R_1^{wc} after obtaining $\mathbf{W}$. Note that the rate expressions in (50a) and (50b) simplify to (10a) and (24a), respectively, when $\mathcal{H}_1 = \{\hat{\mathbf{h}}_1\}$ and $\mathcal{H}_2 = \{\hat{\mathbf{h}}_2\}$, i.e., when $\epsilon_1 = \epsilon_2 = 0$ and the transmitter has perfect channel information regarding both receivers.

Similarly, for the second user we have

$$\begin{aligned}
R_2^{\text{wc}} &\ge \frac{1}{2}\log_2 \left(1 + \frac{\min\limits_{\mathbf{h}_2 \in \mathcal{H}_2} (\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2)^2}{\max\limits_{\mathbf{h}_2 \in \mathcal{H}_2} (\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + \sigma^2} \right) \\
&\qquad + \frac{1}{2}\log_2 \left(\frac{\sigma^2}{\max\limits_{\mathbf{h}_1 \in \mathcal{H}_1} (\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + \sigma^2} \right) \qquad (51a) \\
&\ge \log_2 \frac{\min\limits_{\mathbf{h}_2 \in \mathcal{H}_2} \left|\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2\right| \sigma}{\max\limits_{\substack{\mathbf{h}_1 \in \mathcal{H}_1,\\ \mathbf{h}_2 \in \mathcal{H}_2}} \left((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + \sigma^2\right)^{\frac{1}{2}} \left((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + \sigma^2\right)^{\frac{1}{2}}}. \qquad (51b)
\end{aligned}$$

Next, we combine the rate expressions in (50b) and (51b) using the weights ρ and $1-\rho$, for any $\rho \in [0,1]$, to formulate the robust design problem

$$\underset{\mathbf{W}}{\text{maximize}} \ \ln \frac{\min\limits_{\mathbf{h}_1 \in \mathcal{H}_1} (\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1)^{\rho} \min\limits_{\mathbf{h}_2 \in \mathcal{H}_2} (\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2)^{1-\rho}}{\max\limits_{\mathbf{h}_2 \in \mathcal{H}_2} \left((\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1)^2 + \sigma^2\right)^{\frac{1}{2}} \max\limits_{\mathbf{h}_1 \in \mathcal{H}_1} \left((\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2)^2 + \sigma^2\right)^{\frac{1}{2}}} \tag{52a}$$

$$\text{s.t.} \ \|\mathbf{W}\|_{\mathrm{F}}^2 \le P_{\text{Tot}}, \tag{52b}$$

$$w_{1i}^2 + w_{2i}^2 \le P_i, \quad i = 1, \dots, N. \tag{52c}$$

Problem (52), in turn, can be expressed as

$$\underset{\mathbf{W},z_1,z_2,\delta_1,\delta_2}{\text{maximize}} \quad \ln \frac{z_1^{\rho}\, z_2^{1-\rho}}{(\delta_1^2+\sigma^2)^{\frac{1}{2}}(\delta_2^2+\sigma^2)^{\frac{1}{2}}} \tag{53a}$$

$$\begin{aligned}
\text{s.t.} \quad & \mathbf{h}_1^{\mathrm{T}}\mathbf{w}_1 \geq z_1 \quad \forall \mathbf{h}_1 \in \mathcal{H}_1, && \text{(53b)}\\
& \mathbf{h}_2^{\mathrm{T}}\mathbf{w}_2 \geq z_2 \quad \forall \mathbf{h}_2 \in \mathcal{H}_2, && \text{(53c)}\\
& |\mathbf{h}_2^{\mathrm{T}}\mathbf{w}_1| \leq \delta_1 \quad \forall \mathbf{h}_2 \in \mathcal{H}_2, && \text{(53d)}\\
& |\mathbf{h}_1^{\mathrm{T}}\mathbf{w}_2| \leq \delta_2 \quad \forall \mathbf{h}_1 \in \mathcal{H}_1, && \text{(53e)}\\
& \|\mathbf{W}\|_{\mathrm{F}}^2 \leq P_{\mathrm{Tot}}, && \text{(53f)}\\
& w_{1i}^2 + w_{2i}^2 \leq P_i, \quad i = 1,\ldots,N. && \text{(53g)}
\end{aligned}$$

Utilizing the expressions of the uncertainty sets $\mathcal{H}_1$ and $\mathcal{H}_2$ in (47), the inequalities in (53b), (53c), (53d), and (53e), respectively, can be replaced by

$$\begin{aligned}
& \hat{\mathbf{h}}_1^{\mathrm{T}}\mathbf{w}_1 - \epsilon_1\|\mathbf{w}_1\|_2 \geq z_1, && \text{(54a)}\\
& \hat{\mathbf{h}}_2^{\mathrm{T}}\mathbf{w}_2 - \epsilon_2\|\mathbf{w}_2\|_2 \geq z_2, && \text{(54b)}\\
& |\hat{\mathbf{h}}_2^{\mathrm{T}}\mathbf{w}_1| + \epsilon_2\|\mathbf{w}_1\|_2 \leq \delta_1, && \text{(54c)}\\
& |\hat{\mathbf{h}}_1^{\mathrm{T}}\mathbf{w}_2| + \epsilon_1\|\mathbf{w}_2\|_2 \leq \delta_2. && \text{(54d)}
\end{aligned}$$

Similar to (28), let $f(\delta_1, \delta_2)$ denote the optimal value of the perturbed problem

$$\begin{aligned}
\underset{\mathbf{W},z_1,z_2}{\text{maximize}} \quad & \rho \ln z_1 + (1-\rho)\ln z_2 && \text{(55a)}\\
\text{s.t.} \quad & \text{(54a), (54b), (54c), (54d), (53f), (53g).} && \text{(55b)}
\end{aligned}$$

Then, the robust design problem (53) can be expressed as

$$\underset{\delta_1,\delta_2\geq 0}{\text{maximize}} \quad f(\delta_1,\delta_2) - \frac{1}{2}\ln\left((\delta_1^2+\sigma^2)(\delta_2^2+\sigma^2)\right). \tag{56}$$

Again, we shall refer to (56) as the outer problem, and to (55) as the inner problem. It is clear that the inner problem (55) is convex, and the outer problem (56) is essentially identical to (29). Thus, it can be shown that Propositions 1 and 2 hold for (56) as well. Consequently, (56) can be solved iteratively using Algorithm 1. In each iteration, the subgradient vector $\nabla_{\mathrm{sub}} f(\delta_1,\delta_2)$ is obtained by solving the dual of the inner problem (55). Such a dual problem can be formulated as

$$\underset{\substack{\lambda_1,\lambda_2,\gamma,\boldsymbol{\mu},\\ \boldsymbol{\tau}_1,\boldsymbol{\tau}_2,\chi_1,\chi_2,\\ \boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\nu_1,\nu_2}}{\text{minimize}} \quad \left(\begin{array}{c} \delta_1\lambda_1 + \delta_2\lambda_2 + P_{\mathrm{Tot}}\gamma \\ +\sum_{i=1}^{N}(P_i\mu_i + \tau_{1i} + \tau_{2i}) \\ -\rho\ln\dfrac{\chi_1}{\rho} - (1-\rho)\ln\dfrac{\chi_2}{1-\rho} \end{array}\right) - 1 \tag{57a}$$

$$\begin{aligned}
\text{s.t.} \quad & \chi_1,\chi_2 > 0, \quad |\nu_1| \leq \lambda_1, \quad |\nu_2| \leq \lambda_2, && \text{(57b)}\\
& \|\chi_1\hat{\mathbf{h}}_1 - \boldsymbol{\eta}_1 - \nu_1\hat{\mathbf{h}}_2\|_2 \leq \lambda_1\epsilon_2 + \chi_1\epsilon_1, && \text{(57c)}\\
& \|\chi_2\hat{\mathbf{h}}_2 - \boldsymbol{\eta}_2 - \nu_2\hat{\mathbf{h}}_1\|_2 \leq \lambda_2\epsilon_1 + \chi_2\epsilon_2, && \text{(57d)}\\
& \gamma \geq 0, \quad \mu_i \geq 0, \quad \gamma + \mu_i > 0, && \text{(57e)}\\
& \begin{bmatrix} \tau_{1i} & \eta_{1i} \\ \eta_{1i} & 4(\gamma+\mu_i) \end{bmatrix} \succeq 0, && \text{(57f)}\\
& \begin{bmatrix} \tau_{2i} & \eta_{2i} \\ \eta_{2i} & 4(\gamma+\mu_i) \end{bmatrix} \succeq 0, && \text{(57g)}\\
& \qquad\qquad\qquad i = 1,\ldots,N,
\end{aligned}$$

where λ_1 and λ_2 are the Lagrange multipliers associated with the constraints (54c) and (54d), respectively. Derivation of the dual problem (57) is omitted for brevity. For the special case $\epsilon_1 = \epsilon_2 = 0$, it can be shown that (57) reduces to (35), which is the dual problem with perfect channel information.

C. Robust Precoder Design Subject to Amplitude Constraints

For the case of amplitude constraints, we use the definitions in (49) to obtain the worst-case counterparts of the secrecy rate expressions in (20). Furthermore, the inner problem (55) is modified to

$$\begin{aligned}
\underset{\mathbf{W},z_1,z_2}{\text{maximize}} \quad & \rho \ln z_1 + (1-\rho)\ln z_2 && \text{(58a)}\\
\text{s.t.} \quad & \text{(54a), (54b), (54c), (54d),} && \text{(58b)}\\
& |w_{1i}| + |w_{2i}| \leq A_i, \quad i = 1,\ldots,N, && \text{(58c)}
\end{aligned}$$

and it can be shown that the dual of (58) is

$$\underset{\substack{\lambda_1,\lambda_2,\boldsymbol{\mu},\chi_1,\chi_2,\\ \boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\nu_1,\nu_2}}{\text{minimize}} \quad \left(\begin{array}{c} \delta_1\lambda_1 + \delta_2\lambda_2 + \sum_{i=1}^{N}(A_i\mu_i) \\ -\rho\ln\dfrac{\chi_1}{\rho} - (1-\rho)\ln\dfrac{\chi_2}{1-\rho} \end{array}\right) - 1 \tag{59a}$$

$$\begin{aligned}
\text{s.t.} \quad & \chi_1,\chi_2 > 0, \quad |\nu_1| \leq \lambda_1, \quad |\nu_2| \leq \lambda_2, && \text{(59b)}\\
& \|\chi_1\hat{\mathbf{h}}_1 - \boldsymbol{\eta}_1 - \nu_1\hat{\mathbf{h}}_2\|_2 \leq \lambda_1\epsilon_2 + \chi_1\epsilon_1, && \text{(59c)}\\
& \|\chi_2\hat{\mathbf{h}}_2 - \boldsymbol{\eta}_2 - \nu_2\hat{\mathbf{h}}_1\|_2 \leq \lambda_2\epsilon_1 + \chi_2\epsilon_2, && \text{(59d)}\\
& |\eta_{1i}| \leq \mu_i, \quad |\eta_{2i}| \leq \mu_i, \quad i = 1,\ldots,N. && \text{(59e)}
\end{aligned}$$

The dual problem (59) is convex with second-order cone constraints, and thus it can be efficiently solved using CVX [30]. Then, we proceed with the same steps from the previous subsection and use Algorithm 1 to solve the outer problem (56) and obtain the precoder $\mathbf{W}$.

V. Numerical Examples

In this section, we provide four numerical examples to demonstrate the computational complexity and secrecy performance of the proposed linear precoder.

Example 1: Convergence of Algorithm 1.
In the first example, we investigate the number of iterations required by Algorithm 1 to solve the outer problem (29). Similar to [7, Example 2], we consider the two-user MISO BC-CM with $N = 2$, $\mathbf{h}_1 = [1.5, 0]^{\mathrm{T}}$, $\mathbf{h}_2 = [1.801, 0.871]^{\mathrm{T}}$, and total power constraint $P_{\mathrm{Tot}} = 10$. Furthermore, we impose the per-antenna power constraint $P_i = 6, i \in \{1, 2\}$. We consider the weighted secrecy sum rate corresponding to $\rho = 0.5$, i.e., the two users are assigned equal weights and the secrecy sum rate is maximized. The noise variance σ^2 is equal to 1 at both receivers.

Figure 1 shows the trajectory and convergence of $\varphi(\boldsymbol{\delta}^{(k)})$ using Algorithm 1. We choose the initial point $\boldsymbol{\delta}^{(1)} = (0.1, 0.1)$, or, equivalently, $\boldsymbol{\delta}_{\mathrm{dB}}^{(1)} = (-20\text{ dB}, -20\text{ dB})$, and the subgradient vectors are obtained in each iteration by solving the dual problem (35) using the CVX toolbox [30] in conjunction with the MOSEK solver [31]. Algorithm 1 begins from $\boldsymbol{\delta}_{\mathrm{dB}}^{(1)}$ with a fixed step size $\alpha_{\mathrm{dB}}^{\mathrm{Fix}} = 1$ dB. Upon encountering a peak, which is detected by a reduction in $\varphi(\boldsymbol{\delta}^{(k)})$, the step size is gradually reduced until it becomes 0.2 dB, then the algorithm stops. As can be seen from Figure 1, the algorithm converges after 22 iterations, which is quite reasonable. In the next example, we will show how to reduce the number of

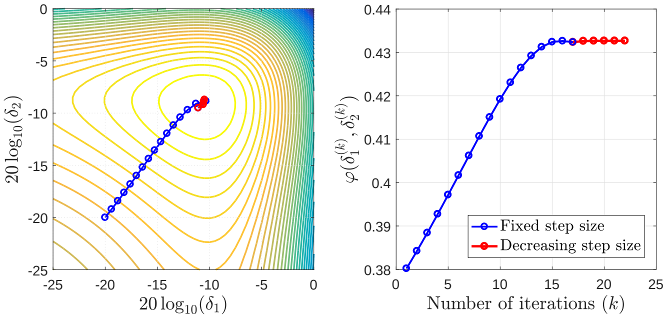

Fig. 1. Trajectory and convergence of $\varphi(\boldsymbol{\delta}^{(k)})$ using Algorithm 1 when applied to the outer problem (29) with $\mathbf{h}_1 = [1.5, 0]^{\mathrm{T}}$, $\mathbf{h}_2 = [1.801, 0.871]^{\mathrm{T}}$, $P_{\text{Tot}} = 10$, $P_1 = P_2 = 6$, and $\rho = 0.5$.

iterations when the outer problem is solved for multiple points corresponding to multiple consecutive values of ρ.

Example 2: Performance comparisons with perfect channel information under total and per-antenna power constraints.

In this example, we use the channel gain vectors $\mathbf{h}_1 = [1.5, 0]^{\mathrm{T}}$ and $\mathbf{h}_2 = [1.801, 0.871]^{\mathrm{T}}$ to demonstrate the achievable secrecy rate regions of the proposed linear precoder under total and per-antenna power constraints. We try two different power levels, particularly, $\{P_{\text{Tot}} = 10, P_1 = P_2 = 6\}$ and $\{P_{\text{Tot}} = 100, P_1 = P_2 = 60\}$. Note that with these levels, both the total and per-antenna power constraints can be active simultaneously. In addition, we include the cases of dominant total power constraint with $P_{\text{Tot}} \in \{10, 100\}$ and dominant per-antenna power constraint with $P_1, P_2 \in \{4, 40\}$. The former case is particularly important because it is the only case for which the secrecy capacity region is precisely known and the boundary points can be calculated using a closed-form expression. This capacity region sets a benchmark that enables us to quantify the loss incurred by using suboptimal linear precoding schemes, and also to validate the algorithm used to obtain the proposed linear precoder.

To plot the secrecy rate regions, we obtain 21 points, i.e., secrecy rate pairs (R_1, R_2), on the boundary of each region by solving the weighted secrecy sum rate maximization problem using $\rho = 0, 0.05, 0.1, \ldots, 1$. For each region, we solve the relevant outer problem using Algorithm 1 to obtain the precoder $\mathbf{W}$, then the rate pairs (R_1, R_2) are calculated by substituting with $\mathbf{W}$ into (10).

Figure 2 depicts the achievable secrecy rate regions of the proposed linear precoder and the zero-forcing (ZF) precoder. The latter is obtained by solving the relevant inner problem using $\delta_1 = \delta_2 = 0$. For the case of dominant total power constraint, we include the secrecy capacity region obtained with the optimal S-DPC scheme [7, Theorem 1] as well as the secrecy rate region of the linear precoder proposed in [9, Corollary 1]. Figure 2 reveals that the secrecy rate regions of the ZF precoder and the linear precoder of [9, Corollary 1] coincide with each other. We can also see that the proposed linear precoder yields better performance, however at the (computational) cost of solving the outer problem.

In order to reduce the total number of iterations used by Algorithm 1 while obtaining the secrecy rate pairs corresponding to $\rho \in \{0, 0.05, 0.1, \ldots, 1\}$, we proceed as follows.

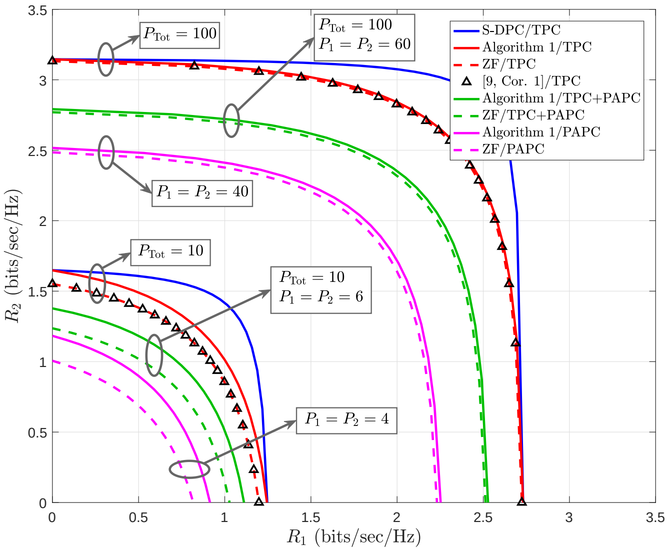

Fig. 2. Achievable secrecy rate regions of the proposed linear precoder (Algorithm 1) and the ZF precoder subject to a total power constraint (TPC), total and per-antenna power constraint (TPC+PAPC), and per-antenna power constraint (PAPC) with $\mathbf{h}_1 = [1.5, 0]^{\mathrm{T}}$ and $\mathbf{h}_2 = [1.801, 0.871]^{\mathrm{T}}$. The secrecy capacity region (S-DPC) and the secrecy rate region of the linear precoder in [9, Corollary 1] are included for the case of total power constraint.

- For the two corner points $\rho = 0$ and $\rho = 1$, the outer problem simplifies to a quasiconvex line search problem, as per Proposition 2. Thus, we solve this problem by performing a bisection search over the interval $[-60\text{ dB}, +20\text{ dB}]$. For each $\rho \in \{0, 1\}$, we obtain a solution with accuracy 0.2 dB in exactly $\left\lceil \log_2 \frac{20-(-60)}{0.2} \right\rceil = 9$ iterations.
- For $\rho = 0.05$, we initiate Algorithm 1 using $\boldsymbol{\delta}_{0.05}^{(1)} = (0.1, 0.1)$. We obtain a solution $\boldsymbol{\delta}_{0.05}^{\star}$ in about 20–30 iterations. Then, for each $\rho \in \{0.1, \ldots, 0.95\}$, we initiate the search using the solution corresponding to the previous value of ρ, that is we choose $\boldsymbol{\delta}_{\rho}^{(1)} = \boldsymbol{\delta}_{\rho-0.05}^{\star}$. For example, with $\rho = 0.1$, the initial point $\boldsymbol{\delta}_{0.1}^{(1)}$ is taken as $\boldsymbol{\delta}_{0.05}^{\star}$.

Figure 3 shows the number of iterations corresponding to three secrecy rate regions from Figure 2. Note that the number of iterations when $\rho \in \{0.1, \ldots, 0.95\}$ is considerably small as compared to $\rho = 0.05$ due to the above procedure.

Example 3: Average performance with perfect channel information.

In this example, we demonstrate the average performance of the proposed precoder using random channel gain vectors under the premise of perfect channel information. The elements of $\mathbf{h}_1$ and $\mathbf{h}_2$ are generated at random (i.i.d. random variables) according to $\mathcal{N}(0, 1)$, and the secrecy rate regions are obtained by averaging over 1000 realizations.

In Figure 4, we plot the secrecy capacity region [7, Theorem 1] along with the secrecy rate region of the proposed linear precoder, subject to a total power constraint specified by $P_{\text{dB}} \triangleq 10\log_{10} P_{\text{Tot}}$. For comparison purposes, we also include the secrecy rate regions of three other linear precoders, namely, the generalized eigenvalue (GEV) precoder, the ZF precoder, and the precoder of [9, Corollary 1]. For the GEV

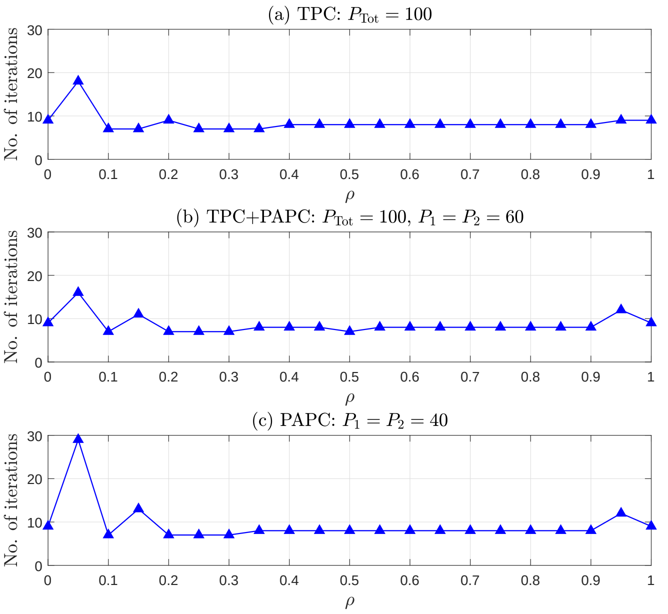

Fig. 3. The number of iterations by Algorithm 1 corresponding to three secrecy rate regions from Figure 2. The corner points at $\rho = 0$ and $\rho = 1$ are the number of iterations with a bisection search.

precoder, the beamformers $\mathbf{w}_{1,\text{GEV}}$ and $\mathbf{w}_{2,\text{GEV}}$ are obtained as follows. Let $\boldsymbol{v}_1$ denote the generalized eigenvector of the matrix pair $(\sigma^2\mathbf{I}_N + P_{\text{Tot}}\mathbf{h}_1\mathbf{h}_1^{\text{T}}, \sigma^2\mathbf{I}_N + P_{\text{Tot}}\mathbf{h}_2\mathbf{h}_2^{\text{T}})$ corresponding to its largest generalized eigenvalue. Then,

$$\mathbf{w}_{1,\text{GEV}} = \sqrt{\rho P_{\text{Tot}}}\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|_2}.$$

Similarly, we have

$$\mathbf{w}_{2,\text{GEV}} = \sqrt{(1-\rho)P_{\text{Tot}}}\frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|_2},$$

where $\boldsymbol{v}_2$ is the generalized eigenvector of the matrix pair $(\sigma^2\mathbf{I}_N + P_{\text{Tot}}\mathbf{h}_2\mathbf{h}_2^{\text{T}}, \sigma^2\mathbf{I}_N + P_{\text{Tot}}\mathbf{h}_1\mathbf{h}_1^{\text{T}})$ corresponding to its largest generalized eigenvalue.

Several interesting conclusions can be drawn from Figure 4. First, we note that the GEV precoder yields better performance than our proposed precoder, especially at low power levels. This is due to the fact that we use the simplified lower bound in (25) as the objective function of the weighted secrecy sum rate maximization problem, rather than the more complex expression in (22d). This, in turn, suggests that GEV is probably a good precoding scheme with low computational complexity when the total power constraint is dominant and channel information is accurately known to the transmitter. Note, however, that there is no counterpart of the GEV scheme for the cases involving per-antenna power or amplitude constraints. We also note from Figure 4 that the proposed precoder yields better performance than the ZF precoder and the precoder from [9, Corollary 1], however at the cost of increased computational complexity. For all the linear precoders, we note that the performance gaps significantly decrease as the number of antennas and/or transmit power increase. Figure 5 shows the average number of iterations to obtain the proposed precoder using Algorithm 1. The figure verifies that the subgradient method is appropriate for tackling the outer problem with low computational complexity and reasonable convergence.

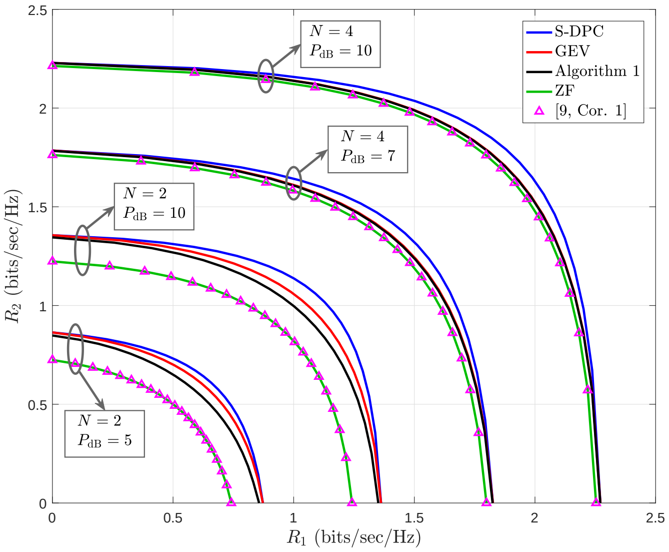

Fig. 4. The secrecy capacity region (S-DPC) along with the secrecy rate regions of the GEV precoder, the proposed linear precoder, the ZF precoder, and the linear precoder in [9, Corollary 1], subject to a total power constraint $P_{\text{dB}} = 10\log_{10} P_{\text{Tot}}$. The number of antennas $N \in \{2, 4\}$.

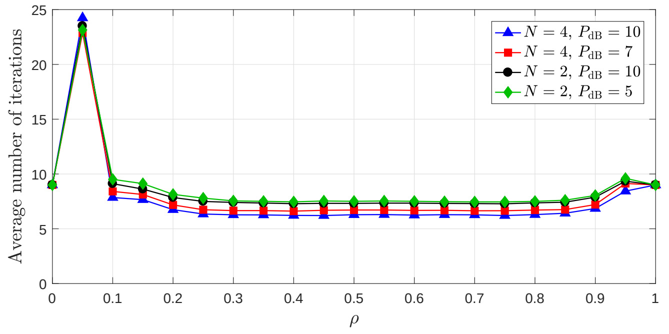

Fig. 5. The average number of iterations for Algorithm 1 corresponding to the secrecy rate regions in Figure 4.

In Figure 6, we plot the achievable secrecy rate regions of the proposed linear precoder, subject to the total power constraint (7), the per-antenna power constraint (11), and the amplitude constraint (13). The secrecy capacity region (for case of total power constraint) and the secrecy rate regions of the ZF precoder (for all constraints) are also included. The power level indicated in the figure specifies the total power constraint in dB, i.e., $P_{\text{dB}} = 10\log_{10} P_{\text{Tot}}$. For comparison purposes, we choose the per-antenna power constraint as $P_i = P_{\text{Tot}}/N$, and the amplitude constraint as $A_i = \sqrt{P_{\text{Tot}}/N}$, for all $i = 1, \ldots, N$. The number of antennas $N = 4$. As expected, the proposed linear precoder outperforms the ZF precoder under all constraints, though at the cost of increased computational complexity.

Example 4: Average performance with imperfect channel information.

Finally, in this example, we illustrate the average performance of the robust linear precoder using random channel realizations. The entires of $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$ are i.i.d. standard

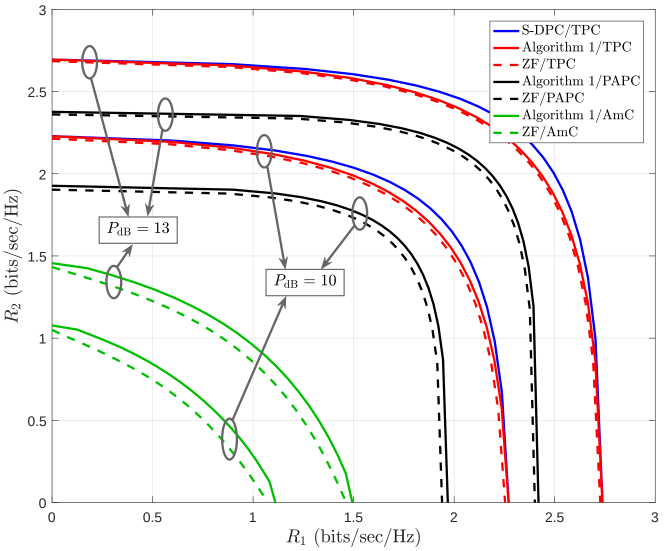

Fig. 6. Achievable secrecy rate regions of the proposed linear precoder and the ZF precoder subject to a total power constraint (TPC), per-antenna power constraint (PAPC) and amplitude constraint (AmC). The number of antennas $N = 4$, and we set $P_{\rm Tot} = 4P_i = 4A_i^2, i = 1,\ldots,4$, and $P_{\rm dB} \triangleq 10\log_{10} P_{\rm Tot}$. The secrecy capacity region (S-DPC) is included for the case of total power constraint.

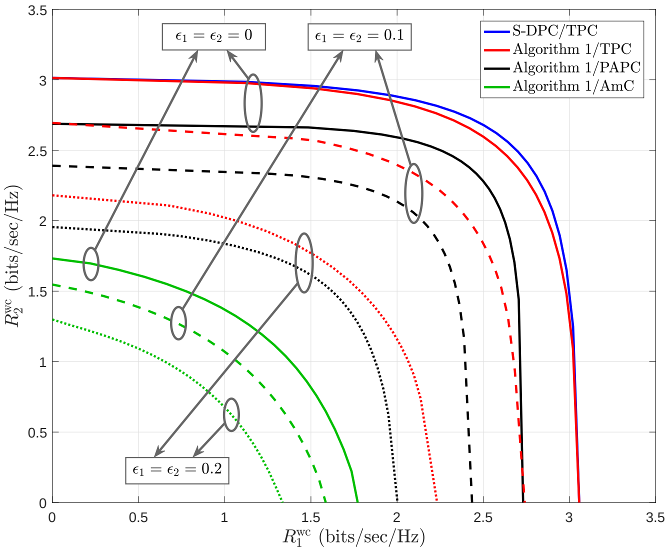

Fig. 7. The worst-case secrecy rate regions with the robust linear precoder under different channel uncertainty levels, $\epsilon_1, \epsilon_2 \in \{0, 0.1, 0.2\}$, subject to a total power constraint (TPC), per-antenna power constraint (PAPC), and amplitude constraint (AmC). We set $N = 4, P_{\rm Tot} = 4P_i = 4A_i^2, i = 1,\ldots,4$, and $10\log_{10} P_{\rm Tot} = 15$ dB. The secrecy capacity region (S-DPC) is included for the case of total power constraint and perfect channel information.

Gaussian random variables, and the secrecy rate regions are averaged over 1000 realizations. In Figure 7, we plot the worst-case secrecy rate regions obtained with the robust precoder considered in Section IV, subject to (7), (11), and (13). Similar to Example 3, we choose $P_i = P_{\rm Tot}/N$ and $A_i = \sqrt{P_{\rm Tot}/N}$, for all $i = 1,\ldots,N$, where $10\log_{10} P_{\rm Tot} = 15$ dB and $N = 4$. The case $\epsilon_1 = \epsilon_2 = 0$ designates perfect channel information, and is included for comparison purposes. As expected, we note from Figure 7 that increased uncertainty levels have negative impact on the worst-case secrecy rate region.

VI. Conclusions

In this paper, we considered the design of linear precoders for the two-user MISO BC-CM subject to total and per-antenna power constraints, and also subject to amplitude constraints. Per-antenna constraints are typically more difficult to handle, but they are essential for modeling hardware limitations in practical systems employing multiple transmit antennas. Although suboptimal, linear precoding is particularly attractive because of low implementation complexity. On the other hand, the optimal S-DPC scheme is difficult to implement, and can be only found via an exhaustive search when per-antenna power constraints are taken into account. Furthermore, the optimal scheme is unknown for the case of amplitude constraints. Therefore, our proposed linear precoding scheme provides a viable solution to an open problem that has not been addressed in the published literature.

We formulated the precoder design problem as a weighted secrecy sum rate maximization problem that is transformed into a more tractable one having only two optimization variables. We proposed a subgradient-based search algorithm to obtain a solution, and characterized the condition under which the solution would be optimal. Our approach is applicable to general convex constraints on the channel input. It is also applicable to the robust design problem when channel uncertainty is taken into account.

We used the total power constraint case, in which the secrecy capacity region is precisely known, to validate our approach and compare the performance of the proposed linear precoder with the optimal S-DPC scheme. The numerical results show a small loss in performance when the SNR is sufficiently high. Compared to the idealistic case of total power constraint and perfect channel information, the results show considerable reduction in the achievable secrecy rate region when per-antenna constraints and channel uncertainty are taken into account.

Appendix A
Proof of Proposition 2

Consider the unit vector $\boldsymbol{u} = [u_1\ u_2]^{\rm T}$, $u_1 \geq 0$, $u_2 \geq 0$, $\|\boldsymbol{u}\|_2 = 1$, and let $\varphi_{\boldsymbol{u}}(t)$, $t \geq 0$, denote the function φ from (29b) with its domain restricted to the line passing through the origin along the direction $\boldsymbol{u}$, i.e.,

$$\begin{aligned}\varphi_{\boldsymbol{u}}(t) &\triangleq \varphi(t\boldsymbol{u}) = \varphi(tu_1, tu_2)\\ &= f_{\boldsymbol{u}}(t) - \frac{1}{2}\left(\ln(u_1^2t^2+\sigma^2) + \ln(u_2^2t^2+\sigma^2)\right), \quad (60)\end{aligned}$$

where $f_{\boldsymbol{u}}(t) \triangleq f(t\boldsymbol{u})$. Our goal here is to prove that, for any $\boldsymbol{u}$, there exists one point $t^\star$ such that $\varphi_{\boldsymbol{u}}(t)$ is nondecreasing for $t \in [0, t^\star]$ and nonincreasing for $t \geq t^\star$, i.e., $\varphi_{\boldsymbol{u}}(t)$ is quasiconcave.

Since $f(\delta_1, \delta_2)$ is concave, its restriction to a line is also concave. As a consequence, $f_{\boldsymbol{u}}(t)$ is continuous and twice differentiable *almost* everywhere, meaning that there are only countably many points where $f''_{\boldsymbol{u}}(t)$ may not exist [33, Chapter 13]. In order to simplify the notation, we will first restrict ourselves to the points at which $f_{\boldsymbol{u}}(t)$ is twice differentiable, then we will see that the extension to all $t > 0$ is straightforward. Differentiating (60) w.r.t. t, we obtain

$$\varphi'_{\boldsymbol{u}}(t) = f'_{\boldsymbol{u}}(t) - \left(\frac{u_1^2 t}{u_1^2 t^2 + \sigma^2} + \frac{u_2^2 t}{u_2^2 t^2 + \sigma^2} \right). \tag{61}$$

Further differentiation yields

$$\varphi''_{\boldsymbol{u}}(t) = f''_{\boldsymbol{u}}(t) + u_1^2 \frac{u_1^2 t^2 - \sigma^2}{(u_1^2 t^2 + \sigma^2)^2} + u_2^2 \frac{u_2^2 t^2 - \sigma^2}{(u_2^2 t^2 + \sigma^2)^2}. \tag{62}$$

Let $t^\star$ denote any point at which $\varphi'_{\boldsymbol{u}}(t) = 0$. Then, we need to show that there is only one such point. Setting $t = t^\star$ and substituting with $\varphi'_{\boldsymbol{u}}(t^\star) = 0$ in (61) yield

$$f'_{\boldsymbol{u}}(t^\star) = \frac{u_1^2 t^\star}{u_1^2 t^{\star 2} + \sigma^2} + \frac{u_2^2 t^\star}{u_2^2 t^{\star 2} + \sigma^2}. \tag{63}$$

Using (62) and (63), $\varphi''_{\boldsymbol{u}}(t^\star)$ can be written as

$$\begin{aligned} \varphi''_{\boldsymbol{u}}(t^\star) = f''_{\boldsymbol{u}}(t^\star) + (f'_{\boldsymbol{u}}(t^\star))^2 - \frac{2 u_1^2 u_2^2 t^{\star 2}}{(u_1^2 t^{\star 2} + \sigma^2)(u_2^2 t^{\star 2} + \sigma^2)} \\ - \frac{u_1^2 \sigma^2}{(u_1^2 t^{\star 2} + \sigma^2)^2} - \frac{u_2^2 \sigma^2}{(u_2^2 t^{\star 2} + \sigma^2)^2}. \end{aligned} \tag{64}$$

Now we will show that the sum $f''_{\boldsymbol{u}}(t^\star) + (f'_{\boldsymbol{u}}(t^\star))^2$ is always nonpositive, and thus $\varphi''_{\boldsymbol{u}}(t^\star)$ is also nonpositive. To do this, we first need to show that $e^{f_{\boldsymbol{u}}(t)}$ is a concave function. Let $G(\delta_1, \delta_2)$ denote the optimal value of the perturbed problem

$$\underset{\mathbf{W}}{\text{maximize}} \quad (\mathbf{h}_1^{\mathrm{T}} \mathbf{w}_1)^{\rho} (\mathbf{h}_2^{\mathrm{T}} \mathbf{w}_2)^{1-\rho} \tag{65a}$$

$$\text{s.t.} \quad |\mathbf{h}_2^{\mathrm{T}} \mathbf{w}_1| \leq \delta_1, \quad |\mathbf{h}_1^{\mathrm{T}} \mathbf{w}_2| \leq \delta_2, \tag{65b}$$

$$\|\mathbf{W}\|_{\mathrm{F}}^2 \leq P_{\mathrm{Tot}}, \tag{65c}$$

$$w_{1i}^2 + w_{2i}^2 \leq P_i, \quad i = 1, \ldots, N. \tag{65d}$$

Since the objective function in (65a) is concave (see [24, Problem 3.16 (f)]), the perturbed problem (65) is convex, and thus $G(\delta_1, \delta_2)$ is a concave function. Next, we note from (28) and (65) that $G(\delta_1, \delta_2) = e^{f(\delta_1, \delta_2)}$. Thus, $G_{\boldsymbol{u}}(t) \triangleq G(t\boldsymbol{u}) = e^{f_{\boldsymbol{u}}(t)}$, and we have

$$G''_{\boldsymbol{u}}(t) = G_{\boldsymbol{u}}(t) \left(f''_{\boldsymbol{u}}(t) + (f'_{\boldsymbol{u}}(t))^2 \right). \tag{66}$$

Since $G_{\boldsymbol{u}}(t)$ is concave, it holds that $G''_{\boldsymbol{u}}(t) \leq 0$ [24, Section 3.1.4]. Furthermore, since $G_{\boldsymbol{u}}(t)$ is nonnegative, we must have

$$f''_{\boldsymbol{u}}(t) + (f'_{\boldsymbol{u}}(t))^2 \leq 0. \tag{67}$$

Thus, $f''_{\boldsymbol{u}}(t^\star) + (f'_{\boldsymbol{u}}(t^\star))^2 \leq 0$ and, consequently, $\varphi''_{\boldsymbol{u}}(t^\star) \leq 0$. The last inequality tells us that $\varphi'_{\boldsymbol{u}}(t)$ can experience zero-crossing only from positive to negative. Since this can happen only once, we conclude that there is only one point $t^\star$ such that

$$\begin{cases} \varphi'_{\boldsymbol{u}}(t) \geq 0 & \text{for} \quad t \leq t^\star, \\ \varphi'_{\boldsymbol{u}}(t) \leq 0 & \text{for} \quad t \geq t^\star. \end{cases}$$

Hence, $\varphi_{\boldsymbol{u}}(t)$ is quasiconcave.

In order to extend the proof to include the points at which $f_{\boldsymbol{u}}(t)$ is non-differentiable, we just need to replace the derivative of $f_{\boldsymbol{u}}(t)$ with any element from its *subdifferential*. Specifically, since $f_{\boldsymbol{u}}(t)$ is concave, it is continuous and has right and left derivatives over the whole interior of its domain (i.e., for all $t > 0$) [34, Theorem 1.6]. Such derivatives are nonincreasing in the sense that, for any $t_2 > t_1 > 0$, we have

$$f'_{\boldsymbol{u}}(t_1^-) \geq f'_{\boldsymbol{u}}(t_1^+) \geq f'_{\boldsymbol{u}}(t_2^-) \geq f'_{\boldsymbol{u}}(t_2^+). \tag{68}$$

Now, at the points where $f'_{\boldsymbol{u}}(t^+) \neq f'_{\boldsymbol{u}}(t^-)$, i.e., $f_{\boldsymbol{u}}(t)$ is non-differentiable, we will allow $f''_{\boldsymbol{u}}(t) \rightarrow -\infty$ and let $f'_{\boldsymbol{u}}(t)$ take any value in the interval $[f'_{\boldsymbol{u}}(t^+), f'_{\boldsymbol{u}}(t^-)]$, which makes (67) hold for all $t > 0$. Thus, $\varphi''_{\boldsymbol{u}}(t^\star)$ is always nonpositive including, possibly, $\varphi''_{\boldsymbol{u}}(t^\star) \rightarrow -\infty$. In other words, $\varphi'_{\boldsymbol{u}}(t^+)$ (or, equivalently, $\varphi'_{\boldsymbol{u}}(t^-)$) can experience zero-crossing only from positive to negative, even if $\varphi'_{\boldsymbol{u}}(t^+)$ has jump discontinuity at the crossing point. Following the same argument for the differentiable case, we conclude that $\varphi_{\boldsymbol{u}}(t)$ is quasiconcave for all $t \geq 0$. ■

REFERENCES

[1] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, pp. 1355–1387, 1975.

[2] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Transactions on Information Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.

[3] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[4] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: The MISOME wiretap channel," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3088–3104, Jul. 2010.

[5] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—Part II: The MIMOME wiretap channel," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5515–5532, Nov. 2010.

[6] R. Liu, I. Maric, P. Spasojevic, and R. D. Yates, "Discrete memoryless interference and broadcast channels with confidential messages: Secrecy rate regions," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2493–2507, Jun. 2008.

[7] R. Liu and H. Poor, "Secrecy capacity region of a multiple-antenna Gaussian broadcast channel with confidential messages," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1235–1249, Mar. 2009.

[8] R. Liu, T. Liu, H. V. Poor, and S. Shamai, "Multiple-input multiple-output Gaussian broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4215–4227, Sept. 2010.

[9] S. A. A. Fakoorian and A. L. Swindlehurst, "On the optimality of linear precoding for secrecy in the MIMO broadcast channel," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1701–1713, Sept. 2013.

[10] D. Park, "Weighted sum rate maximization of MIMO broadcast and interference channels with confidential messages," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 1742–1753, Mar. 2016.

[11] F. Boccardi and H. Huang, "Optimum power allocation for the MIMO-BC zero-forcing precoder with per-antenna power constraints," in *2006 40th Annual Conference on Information Sciences and Systems*, Mar. 2006, p. 504.

[12] S. Shi, M. Schubert, and H. Boche, "Per-antenna power constrained rate optimization for multiuser MIMO systems," in *2008 International ITG Workshop on Smart Antennas*, Feb. 2008, pp. 270–277.

[13] P. L. Cao, T. J. Oechtering, R. F. Schaefer, and M. Skoglund, "Optimal transmit strategy for MISO channels with joint sum and per-antenna power constraints," *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4296–4306, Aug. 2016.

[14] P. L. Cao, T. J. Oechtering, and M. Skoglund, "Optimal transmission with per-antenna power constraints for multiantenna bidirectional broadcast channels," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jul. 2016, pp. 1–5.

[15] Q. Li, M. Hong, H. T. Wai, Y. F. Liu, W. K. Ma, and Z. Q. Luo, "Transmit solutions for MIMO wiretap channels using alternating optimization," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1714–1727, Sept. 2013.

[16] Z. Ghassemlooy, S. Arnon, M. Uysal, Z. Xu, and J. Cheng, "Emerging optical wireless communications – Advances and challenges," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 9, pp. 1738–1749, Sept. 2015.

[17] J. G. Smith, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Journal of Information and Control*, vol. 18, pp. 203–219, 1971.

[18] A. Lapidoth, S. Moser, and M. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4449–4461, Oct. 2009.

[19] A. A. Farid and S. Hranilovic, "Capacity bounds for wireless optical intensity channels with Gaussian noise," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6066–6077, Dec. 2010.

[20] H. Ma, L. Lampe, and S. Hranilovic, "Coordinated broadcasting for multiuser indoor visible light communication systems," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3313–3324, Sept. 2015.

[21] A. Mostafa and L. Lampe, "Optimal and robust beamforming for secure transmission in MISO visible-light communication links," *IEEE Transactions on Signal Processing*, vol. 64, no. 24, pp. 6501–6516, Dec. 2016.

[22] A. Mostafa and L. Lampe, "Physical-layer security for MISO visible light communication channels," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 9, pp. 1806–1818, Sept. 2015.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2006.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009.

[25] N. Z. Shor, *Minimization Methods for Non-differentiable Functions*. Springer Series in Computational Mathematics. Springer, 1985.

[26] S. Boyd, L. Xiao, and A. Mutapcic, *Subgradient methods*. Notes for EE392o, Stanford University, Autumn, 2003.

[27] G. Calafiore and L. El Ghaoui, *Optimization Models*. Cambridge University Press, 2014.

[28] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.

[29] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

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

[31] M. ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).*, 2015. [Online]. Available: http://docs.mosek.com/7.1/toolbox/index.html

[32] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[33] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Springer, 2009.

[34] J. V. Tiel, *Convex Analysis: An Introductory Text*. John Wiley & Sons, 1984.

Ayman Mostafa (S'08 - M'17) received the B.Sc. degree (with honours) in electrical engineering from Alexandria University, Alexandria, Egypt, in 2006, the M.A.Sc. degree in electrical engineering from McMaster University, Hamilton, ON, Canada, in 2012, and the Ph.D. degree in electrical engineering from The University of British Columbia, Vancouver, BC, Canada, in 2017. He is currently a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada. His research interests are in the areas of communication and information theory, physical-layer security, and signal processing for communications.

Lutz Lampe (M'02 - SM'08) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the University of Erlangen, Germany, in 1998 and 2002, respectively. Since 2003, he has been with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada, where he is a Full Professor. His research interests are broadly in theory and application of wireless, power line, optical wireless and optical fibre communications. Dr. Lampe is currently an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE COMMUNICATIONS LETTERS, and the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He was a (co-)recipient of a number of best paper awards, including awards at the 2006 IEEE International Conference on Ultra-Wideband (ICUWB), the 2010 IEEE International Communications Conference (ICC), and the 2011 and 2017 IEEE International Conference on Power Line Communications and Its Applications (ISPLC). He was the General (Co-)Chair for the 2005 IEEE ISPLC, the 2009 IEEE ICUWB and the 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm). He is a co-edior of the book "Power Line Communications: Principles, Standards and Applications from Multimedia to Smart Grid," published by John Wiley & Sons in its 2nd edition in 2016.