Signaling Design for MIMO-NOMA With Different Security Requirements

Yue Qi, *Graduate Student Member, IEEE*, and Mojtaba Vaezi, *Senior Member, IEEE*

***Abstract*—Signaling design for secure transmission in two-user multiple-input multiple-output (MIMO) non-orthogonal multiple access (NOMA) networks with different security requirements is investigated. A base station broadcasts multicast data to all users and unicast data and confidential data targeted to certain users. We categorize the above channel into three communication scenarios depending on the security requirements. The associated problem in each scenario is nonconvex. We propose a unified approach, called the power splitting scheme, for optimizing the rate equations corresponding to each scenario. The proposed method converts the optimization of the secure MIMO-NOMA channel into a set of simpler problems, namely multicast, point-to-point, and wiretap MIMO problems, corresponding to the three basic messages: multicast, private/unicast, and confidential messages. We then leverage existing solutions to design signaling (covariance matrix) for the above problems such that the messages are transmitted with high security and reliability. Numerical results illustrate the efficacy of the proposed covariance matrix (linear precoding and power allocation) design. In the case of no multicast messages, we also reformulate the nonconvex problem into weighted sum rate (WSR) maximization problems by applying the block successive maximization method and generalizing the zero duality gap. The two methods have their advantages and limitations. Power splitting is a general tool that can be applied to the MIMO-NOMA with any combination of the three messages (multicast, private, and confidential) whereas the WSR maximization shows greater potential for secure MIMO-NOMA communication without multicasting. In such cases, the WSR maximization provides a slightly better rate than the power splitting method.**

***Index Terms*—MIMO-NOMA, broadcast channel, physical layer security, power splitting, weighted sum rate, wiretap, multicast, unicast.**

I. INTRODUCTION

THE unprecedented wave of emerging devices has dramatically increased the requirements and challenges of resource allocation and spectrum utilization. To fulfill the demands, non-orthogonal multiple access (NOMA) at the physical (PHY) layer is a promising technique [3], [4] that has attracted remarkable attention both in academia and industry.

Manuscript received September 28, 2020; revised June 29, 2021, October 25, 2021, and February 14, 2022; accepted February 28, 2022. Date of publication March 7, 2022; date of current version March 22, 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ali Tajer. This work was presented in part at IEEE Global Communications Conference [DOI: 10.1109/GLOBECOM42002.2020.9348124] and the IEEE International Symposium on Dynamic Spectrum Access Networks [DOI: 10.1109/DySPAN53946.2021.9677232]. *(Corresponding author: Yue Qi.)*

The authors are with the Villanova University, Villanova, PA 19085-1603 USA (e-mail: yqi@villanova.edu; mvaezi@villanova.edu).

Digital Object Identifier 10.1109/TSP.2022.3156915

While several code-domain uplink NOMA schemes are developed in the literature [4], [5], downlink NOMA is based on well-known information-theoretic techniques for the broadcast channel (BC) [6]. Then, in the single-input single-output (SISO) NOMA, superposition coding (SC) at the transmitter and successive interference cancellation (SIC) at the receiver give the optimal strategy. Hence, a large body of work has assumed NOMA to be equivalent to SC-SIC, and applied SC-SIC to multiple-input multiple-output (MIMO) channels [7]–[10]. However, it is known that SC-SIC is not capacity-achieving in the MIMO-BC and *dirty-paper coding* (DPC) is the optimal strategy [11]–[13]. Similarly, in MIMO-NOMA with PHY layer security, SC-SIC cannot achieve secure capacity, and secret DPC (S-DPC) is the optimal solution [14], [15]. In this paper, NOMA is defined broadly and refers to any technique that allows simultaneous transmission over the same resources [12], i.e., concurrent non-orthogonal transmission. That is, MIMO-NOMA is equivalent to the MIMO-BC.

A. MIMO-NOMA With Secrecy

Today, there is a trend to merge multiple services in one transmission. This is referred to as *PHY layer service integration* [16]. Integrated services usually include three fundamental services: multicast, unicast, and confidential services, which can be realized by common, private/individual, and confidential messages, respectively. Especially, secure transmission of confidential messages requires PHY layer security which has been introduced as additional protection for secure transmission [17].

This work is concerned with different security requirements for two-user MIMO-NOMA networks, in which three different types of messages can be transmitted:

- *Common message M_0 [18]:* a common message is transmitted in such a way that all users can decode it. For example, the base station (BS) broadcasts daily news or amber alerts to all online users.
- *Private message M_p [6]:* a private or unicast/individual message is a message intended for a specific user. For instance, the BS provides targeted advertisements and recommended videos that are available only to interested users. This message is not encoded securely, and as such, it can be decoded by other users.
- *Confidential message M_c [19]:* a confidential message is similar to a private message but is to be kept secret from other users. For example, personal email accounts access and online banking transactions. Here, encoding is such that the message cannot be decoded by others.

TABLE I
A SUMMARY OF COMMUNICATION SCENARIOS WITH DIFFERENT COMBINATIONS OF COMMON, PRIVATE, AND CONFIDENTIAL MESSAGES

	Communication scenarios	M_0	M_1	M_2	Capacity region	Signaling schemes
OMA	Multicasting	Public	–	–	[18]	Heuristic precoding [27], closed-form [1]
	P2P MIMO	–	Private	–	[28]	SVD and WF [28]
	Wiretap	–	Confidential	–	[19]	GSVD [29], AOWF [30], RM [31]
NOMA	Two private	–	Private	Private	[6], [32]–[34]	GSVD [35], **this work**
	Private and confidential	–	Confidential	Private	[23]	**This work**
	Two confidential	–	Confidential	Confidential	[14]	GSVD [36], BSMM [37], PS [38]
	Common and one confidential	Public	Confidential	–	[39]	GSVD [40], RM [41]
	Scenario A	Public	Private	Private	[11], [20], [21]	**This work**
	Scenario B	Public	Confidential	Private	[22], [23]	**This work**
	Scenario C	Public	Confidential	Confidential	[15], [24]	**This work**

Early information-theoretic works [11], [15], [20]–[23] have established the capacity regions of two-user MIMO-BC with different security requirements. These include the MIMO-BC with one common and two independent private messages [11], [20], [21], the MIMO-BC with private, confidential, and common messages [22], [23], and the MIMO-BC with one common and two confidential messages [15], [24]. However, their primary purpose is to derive capacity regions or to construct coding strategies that achieve certain rate regions. The solutions are based on DPC or S-DPC and usually are given as a union over all possible transmit covariance matrices satisfying certain power constraints. Implementation of DPC requires sophisticated random coding [25], and finding practical dirty paper codes close to the capacity is not easy [26]. Linear precoding is a popular alternative to simplify the transmission design [13], [26].

The two-user MIMO-NOMA can be classified into three communication scenarios with different security requirements as shown in Fig 1. The classification is mainly based on the well-established information-theoretic results:

- Scenario A (no security): two independent private messages M_{1p} and M_{2p} (one for each user) and a common message M_0 for both users are ordered [11], [20], [21]. In this case, we have a MIMO-NOMA with common and two private messages (M_0, M_{1p}, M_{2p});
- Scenario B (security for one user): a confidential message M_{1c} for user 1, a private message M_{2p} for user 2, and one common message M_0 for both users are ordered [23]. Then, a MIMO-NOMA with common, private, and confidential messages (M_0, M_{1c}, M_{2p}) is formed;
- Scenario C (security for both users): two confidential messages M_{1c} and M_{2c} (one for each user), and a common message M_0 for both users are needed [15], [24]. In this case, a MIMO-NOMA with common and confidential messages (M_0, M_{1c}, M_{2c}) is obtained.

The three scenarios overall cover nine problems, or communication scenarios (see Table I). The combinations of different types of messages are also named integrated services [16].

B. Motivation and Related Problems

While the capacity regions of the three different messages are characterized, it is still unknown how to identify optimal or implementation-efficient solutions to achieve those regions. Thus, this paper is motivated by the following question: How can we maximize the secrecy rate for the MIMO-NOMA with different security requirements in an acceptable computational complexity?

The state-of-the-art includes solutions only for some special combinations of the messages and the orthogonal multiple access (OMA) case in which only one message, out of the three messages mentioned earlier, is transmitted. These are summarized in Table I, and some are highlighted below.

- *Two private messages [6], [32], [33]:* When there is no common message in Scenario A, the problem reduces to the MIMO-BC and DPC gives the capacity region. Alternatively, the multiple access channel (MAC) to BC duality [32] can be applied to iteratively achieve the capacity [34]. Also, an analytic linear precoding scheme based on *generalized singular value decomposition* (GSVD) is designed for the special case where the two users are equipped with the same number of antennas [35].
- *One confidential message [19]:* When there is neither a common nor private message in Scenario B, the system reduces to the MIMO wiretap channel [19]. Various linear precoding schemes, including GSVD [29], alternating optimization and water filling (AOWF) [30], and rotation modeling (RM) [31] are known for this problem.
- *Two confidential messages [14]:* When M_0 is empty in Scenario C, the problem reduces to the MIMO-BC with two confidential messages. It is proven in [14, Theorem 1] that both users can reach their respective maximum secrecy rates simultaneously by S-DPC. Low-complexity approaches, such as GSVD [36], weighted sum rate (WSR) maximization with block successive maximization method (BSMM) [37], and power splitting (PS) method [38] are developed. We generalize the PS into a more general case.
- *Common and confidential messages [39]:* Different linear precoding schemes, including GSVD-based precoding [40] and RM with random search [41] are known in this case.
- *Only one common message [18], [27]:* If there is only a common message and no individual messages to be transmitted, the system becomes a multicast channel. A heuristic precoding with iterations is investigated in [27], and an analytical solution with a convex tool is given in [1].

However, these problems are all special cases of the three general scenarios mentioned earlier. Signaling designs for the general cases are still unknown in general.

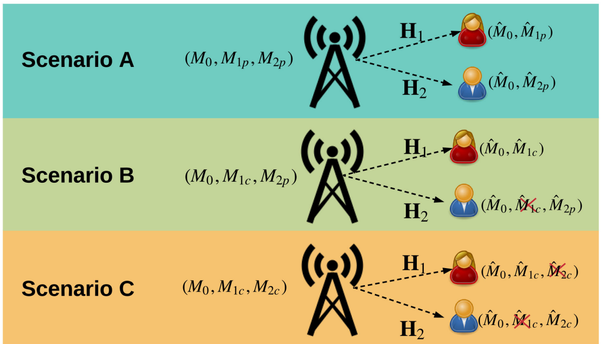


Fig. 1. Communication scenarios with different combinations of security requirements based on the information-theoretic results. Consider three communication scenarios in which the BS sends different combinations of the three messages.

C. Contributions and Organization

As illustrated, the problems listed in Table I are all related to the three scenarios shown in Fig. 1. Nonetheless, there are no solutions for the general cases. In this paper, we propose a new solution, named the power splitting method, which applies to all of those problems. This method decomposes the secure MIMO-NOMA channel into point-to-point (P2P), wiretap, and multicasting MIMO channels. Then, we design one algorithm that can be used in all problems in Table I to approach their corresponding capacity regions. The main contributions of this paper can be summarized as follows:

- We first split the total power among the three messages and then reformulate the secrecy capacity optimization problems into three sub-problems. Particularly, Scenario A (only private massages) is decomposed into two P2P MIMO channels; Scenario B (private and confidential messages) is decomposed into one P2P MIMO and one wiretap channel; and, Scenario C (only confidential messages) is decomposed into two wiretap channels.
- Linear precoder and power allocation matrices are designed for private and confidential messages by extending the analytical solution of the P2P MIMO problem and the numerical solution of the wiretap channel to the MIMO-NOMA with different secrecy scenarios. For multicasting, we use a combination of analytical solutions and a numerical solution based on a convex tool. Finally, we propose an algorithm for all different secrecy scenarios.
- When there is no common message, a WSR maximization is formulated in all scenarios. We prove that the WSR problem has zero duality gap in all scenarios, and the KKT conditions are necessary for the optimal solutions. Besides, we derive and generalize an iterative algorithm for all scenarios by applying the BSMM [37], [42]. Especially, in Scenario A, we provide an alternative solution that directly optimizes the WSR of the DPC region with BSMM instead of applying MAC-BC duality.

One main benefit of the proposed signaling design (power splitting, linear precoding, and power allocation) is its ability to be generalized to more complicated scenarios, e.g., when there are more than two users.

The remainder of this paper is organized as follows. In the next section, we discuss the channel model and formulate the problems for the three scenarios. We introduce a power splitting method to all scenarios in Section III-A, and a signaling design for each in Section III-B. For the subcases without common messages, we generalize a WSR based on BSMM for all scenarios in Section IV. We then present numerical results in Section V and conclude the paper in Section VI.

Notations: $\mathrm{tr}(\cdot)$ and $(\cdot)^T$ denote trace and transpose of matrices. $E\{\cdot\}$ denotes expectation. $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ represents diagonal matrix with elements $\lambda_1, \ldots, \lambda_n$. $\mathbf{Q} \succcurlyeq \mathbf{0}$ represents that $\mathbf{Q}$ is a positive semidefinite matrix. $[x]^+$ gives the max value of x and 0. $\mathbf{I}$ is an identity matrix.

II. System Model

Considering a two-user MIMO-NOMA network. A BS equipped with n_t antennas simultaneously serves two users, in which user 1 and user 2 are equipped with n_1 and n_2 antennas, respectively. The transmitted signal to user 1 and user 2 share the same time slot and frequency.

The received signals at user 1 and user 2 are given by

$$\mathbf{y}_1 = \mathbf{H}_1 \mathbf{x} + \mathbf{w}_1, \tag{1a}$$

$$\mathbf{y}_2 = \mathbf{H}_2 \mathbf{x} + \mathbf{w}_2, \tag{1b}$$

in which $\mathbf{H}_1 \in \mathbb{R}^{n_1 \times n_t}$ and $\mathbf{H}_2 \in \mathbb{R}^{n_2 \times n_t}$ are the channel matrices for user 1 and user 2, respectively. The elements of the channels are drawn from independent and identically distributed (i.i.d.) Gaussian distributions. $\mathbf{w}_1 \in \mathbb{R}^{n_1 \times 1}$ and $\mathbf{w}_2 \in \mathbb{R}^{n_2 \times 1}$ are i.i.d. Gaussian noise vectors whose elements are zero mean and unit variance. The input $\mathbf{x} \in \mathbb{R}^{n_t \times 1}$ is a vector consisting of three components

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_1 + \mathbf{x}_2, \tag{2}$$

where $\mathbf{x}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$, $k = 0, 1, 2$, is the input corresponding to two kinds of services: the multicast message M_0 and secure messages (private M_p and/or confidential M_c) of user 1 and user 2, in which $\mathbf{Q}_k \succcurlyeq \mathbf{0}$, is the covariance matrix corresponding to $\mathbf{x}_k$. The channel input is subject to an average total power constraint

$$\mathrm{tr}(\mathbb{E}\{\mathbf{x}\mathbf{x}^T\}) = \mathrm{tr}(\mathbf{Q}_0 + \mathbf{Q}_1 + \mathbf{Q}_2) \leq P. \tag{3}$$

We denote R_0, R_{jp}, and R_{jc}, $j = 1, 2$, as the achievable rates associated with the multicast, private, and confidential messages transmitted by the corresponding user j, respectively.

In the following, we provide the achievable rate region for each scenario.

A. Scenario A (One Common and Two Private Messages)

The DPC rate region of the MIMO-NOMA with common and two private messages is realized by [11], [20],

$$R_A(P) = \mathrm{conv}\left\{ \mathcal{R}_{12}^{\mathrm{DPC}} \cup \mathcal{R}_{21}^{\mathrm{DPC}} \right\} \tag{4}$$

in which conv is the convex hull operator. $\mathcal{R}_{12}^{\mathrm{DPC}}$ consists of all triples (R_{1p}, R_{2p}, R_0) satisfying

$$R_0 \leq \min(R_{01}, R_{02}), \tag{5a}$$

$$R_{1p} \leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{Q}_1 \mathbf{H}_1^T|, \tag{5b}$$

$$R_{2p} \leq \frac{1}{2} \log |\mathbf{I} + (\mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1 \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{Q}_2 \mathbf{H}_2^T| \tag{5c}$$

where

$$R_{0j} \triangleq \frac{1}{2} \log \left| \mathbf{I} + \frac{\mathbf{H}_j \mathbf{Q}_0 \mathbf{H}_j^T}{(\mathbf{I} + \mathbf{H}_j(\mathbf{Q}_1 + \mathbf{Q}_2)\mathbf{H}_j^T)} \right|, j = 1, 2 \tag{6}$$

with the total power constraint (3). $\mathcal{R}_{21}^{\mathrm{DPC}}$ is obtained from $\mathcal{R}_{12}^{\mathrm{DPC}}$ by swapping the subscripts 1 and 2 corresponding to different DPC encoding orders. When each user has a single antenna, the problem can be transferred to a linear semi-definite convex optimization [11, Section III], but the MIMO case is in general still unknown. Without the common message, the capacity of the MIMO BC is given in [32], [33].

B. Scenario B (Common, Private, and Confidential Messages)

In this scenario, only user 1 requires a confidential message. The secrecy capacity region $R_B(P)$ under a total power constraint (3) is given by a set of rate triples (R_{1c}, R_{2p}, R_0) satisfying [23, Theorem 2]

$$R_0 \leq \min(R_{01}, R_{02}), \tag{7a}$$

$$R_{1c} \leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{Q}_1 \mathbf{H}_1^T| - \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1 \mathbf{H}_2^T|, \tag{7b}$$

$$R_{2p} \leq \frac{1}{2} \log |\mathbf{I} + (\mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1 \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{Q}_2 \mathbf{H}_2^T|. \tag{7c}$$

The entire secrecy capacity region is achieved using DPC to cancel out the signal of private M_{2p} at user 2, the other variant, i.e., DPC against M_{1c} at user 1, is unnecessary. This is different from Scenario A for which the capacity region is exhausted by taking the convex hull of both variants ($\mathcal{R}_{21}^{\mathrm{DPC}}$ and $\mathcal{R}_{12}^{\mathrm{DPC}}$).

C. Scenario C (One Common and Two Confidential Messages)

The secrecy capacity region $R_C(P)$ of the MIMO-BC with one common and two confidential messages under the average total power constraint (3) can be expressed as [15, Theorem 2]

$$R_0 \leq \min(R_{01}, R_{02}), \tag{8a}$$

$$R_{1c} \leq \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{Q}_1 \mathbf{H}_1^T| - \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1 \mathbf{H}_2^T|, \tag{8b}$$

$$R_{2c} \leq \frac{1}{2} \log |\mathbf{I} + (\mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1 \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{Q}_2 \mathbf{H}_2^T| - \frac{1}{2} \log |\mathbf{I} + (\mathbf{I} + \mathbf{H}_1 \mathbf{Q}_1 \mathbf{H}_1^T)^{-1} \mathbf{H}_1 \mathbf{Q}_2 \mathbf{H}_1^T|, \tag{8c}$$

The secrecy capacity region is characterized by S-DPC [15], [24].[1] In this scenario, both users' transmissions are secret from each other. User 1 with confidential messages M_{1c} treats user 2 as an eavesdropper, and vice versa.

[1]The S-DPC can assure security between the two users because a precoding matrix is selected such that it satisfies two goals [14, Remark 5]. First, it helps to cancel the precoding signal representing message M_{2c}, so that M_{1c} can be served with an interference-free legitimate user channel. Second, it boosts the secrecy for message M_{2c} by causing interference (artificial noise) to user 1. In other words, user 1 can remove the interference of user 2 but is not able to decode the message of user 2.

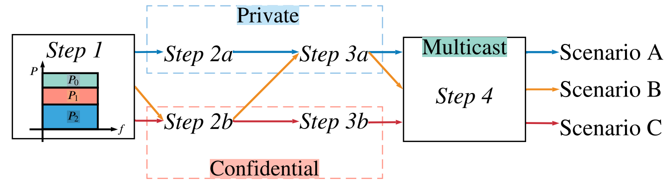


Fig. 2. System structure of power splitting method for different security scenarios.

The border of the secrecy capacity regions in (5), (7), (8) can be obtained by an exhaustive search over the set of all possible input covariance matrices. However, the complexity of such methods is prohibitive for practical implementations, which motivates us to develop a simpler signaling scheme. The covariance matrices achieving the capacity regions are not known in general due to the non-convexity.

III. Power Splitting Method for MIMO-NOMA in All Scenarios

To introduce a new simpler and faster solution, we split the total power for three messages in each scenario. Then, we decouple the MIMO-NOMA channel of all secrecy scenarios into three different problems and solve them separately.

A. Decomposing Secure MIMO-NOMA Into Simpler Channels

We decompose the MIMO-NOMA into different problems in this section. The structure of our decomposition of the MIMO-NOMA into different problems is shown in Fig. 2. Due to some overlapping, such as the privacy part in Scenario A and Scenario B, the confidentiality part in Scenario B and Scenario C, we start with *Step 1* to split the power between user 1 and user 2 for different usages. Then, *Step 2a* and *Step 2b* are for user 1 with private messages in Scenario A and confidentiality in Scenario B, respectively. *Step 3a* and *Step 3b* are for user 2 with private messages in Scenario A, and confidentiality in Scenario C, correspondingly. Lastly, *Step 4* is for common message in all scenarios.

Step 1: Introducing power splitting factors $\alpha_k \in [0, 1]$, $\sum_k \alpha_k = 1$, $k = 0, 1, 2$, we dedicate a fraction α_1 of the total power to user 1, a fraction $\alpha_2 \in [0, \alpha_1]$ to user 2, and allocate the remaining power to the common message M_0 for both users ($P_0 = \alpha_0 P = (1 - \alpha_1 - \alpha_2)P$). The optimal solution uses total power throughout the paper.

Step 2a: We design the secure precoding for user 1 with private message M_{1p} in (5b) for Scenario A. Since the rate $R_{1p}(\alpha_1)$ of user 1 is only controlled by the covariance matrix $\mathbf{Q}_1$ under the power constraint P_1, the interference-free link can be seen as a P2P MIMO with power P_1, which is

$$R_{1p}(\alpha_1) = \max_{\mathbf{Q}_1 \succeq \mathbf{0}} \frac{1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{Q}_1 \mathbf{H}_1^T| \tag{9a}$$

$$\text{s.t.} \quad \mathrm{tr}(\mathbf{Q}_1) \leq P_1 = \alpha_1 P. \tag{9b}$$

The solution $\mathbf{Q}_1^*$ is obtained analytically through singular value decomposition (SVD) and water filling (WF) [28].

Step 2b: In Scenario B and Scenario C, we design secure precoding for user 1 with confidential messages M_{1c} while treating the second user as an eavesdropper. Because covariance matrix $\mathbf{Q}_1$ is the only variable in (7b) and (8b), the problem can be seen as a wiretap channel under a transmit power P_1, which is

$$R_{1c}(\alpha_1) = \max_{\mathbf{Q}_1 \succeq \mathbf{0}} \frac{1}{2} \log \frac{|\mathbf{I} + \mathbf{H}_1\mathbf{Q}_1\mathbf{H}_1^T|}{|\mathbf{I} + \mathbf{H}_2\mathbf{Q}_1\mathbf{H}_2^T|}, \tag{10a}$$

$$\text{s.t.} \quad \mathrm{tr}(\mathbf{Q}_1) \leq P_1 = \alpha P. \tag{10b}$$

This problem is now the well-known MIMO wiretap channel [31], and standard MIMO wiretap solutions can be applied to obtain $\mathbf{Q}_1^*$.

Step 3a: To maximize the secrecy rate $R_{2p}(\alpha_2)$ for user 2, we apply $\mathbf{Q}_1^*$ obtained in *Step 2a* or *Step 2b* to (5c) and (7c) in Scenario A and Scenario B, respectively. Thus, (5c) or (7c) can be represented as

$$R_{2p}(\alpha_2) = \max_{\mathbf{Q}_2 \succeq \mathbf{0}} \frac{1}{2} \log |\mathbf{I} + (\mathbf{I} + \mathbf{H}_2\mathbf{Q}_1^*\mathbf{H}_2^T)^{-1}\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T| \tag{11a}$$

$$\text{s.t.} \quad \mathrm{tr}(\mathbf{Q}_2) \leq P_2 = \alpha_2 P. \tag{11b}$$

Since $\mathbf{Q}_1^*$ is already given, in the following we show that the above problem can be seen as a P2P MIMO problem under power P_2.

Theorem 1: The optimization problem in (11) with interference from user 1 can be converted to the optimization of a standard P2P MIMO channel

$$\dot{\mathbf{H}}_2 \triangleq \mathbf{B}^{-\frac{1}{2}}\mathbf{C}^T\mathbf{H}_2, \tag{12}$$

in which $\mathbf{B}$ and $\mathbf{C}$ are the eigenvalues and eigenvectors of $\mathbf{I} + \mathbf{H}_2\mathbf{Q}_1^*\mathbf{H}_2^T$.

Proof: Define

$$\mathbf{\Sigma} \triangleq \mathbf{I} + \mathbf{H}_2\mathbf{Q}_1^*\mathbf{H}_2^T = \mathbf{C}\mathbf{B}\mathbf{C}^T. \tag{13}$$

Then, the secrecy rate for user 2 can be written as

$$\begin{aligned} R_{2p}(\alpha_2) &= \max_{\mathbf{Q}_2 \succeq \mathbf{0}} \frac{1}{2} \log |\mathbf{I} + \mathbf{\Sigma}^{-1}\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T| \\ &= \max_{\mathbf{Q}_2 \succeq \mathbf{0}} \frac{1}{2} \log |\mathbf{I} + \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T| \\ &\overset{(a)}{=} \max_{\mathbf{Q}_2 \succeq \mathbf{0}} \frac{1}{2} \log |\mathbf{I} + \mathbf{B}^{-\frac{1}{2}}\mathbf{C}^T\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T\mathbf{C}\mathbf{B}^{-\frac{1}{2}}| \\ &= \max_{\mathbf{Q}_2 \succeq \mathbf{0}} \frac{1}{2} \log |\mathbf{I} + \dot{\mathbf{H}}_2\mathbf{Q}_2\dot{\mathbf{H}}_2^T|, \end{aligned} \tag{14}$$

in which (a) holds because of Sylvester's determinant theorem, i.e., $\det(\mathbf{I} + \mathbf{X}\mathbf{Y}) = \det(\mathbf{I} + \mathbf{Y}\mathbf{X})$ where $\mathbf{X} = \mathbf{C}\mathbf{B}^{-\frac{1}{2}}$ and $\mathbf{Y} = \mathbf{B}^{-\frac{1}{2}}\mathbf{C}^T\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T$. $\mathbf{C}$ is orthogonal, i.e., $\mathbf{C}^{-1} = \mathbf{C}^T$, and $\mathbf{B}$ is a diagonal matrix. □

In view of (14), the problem in (11) becomes the standard P2P MIMO without interference over a modified channel. The solution $\mathbf{Q}_2^*$ can be obtained the same as *Step 2a*.

Step 3b: To maximize the secrecy rate $R_{2c}(\alpha_2)$ for user 2, we apply $\mathbf{Q}_1^*$ obtained in *Step 2b* to (8c) for Scenario C. Thus, (8c) can be represented as

$$R_{2c}(\alpha_2) = \max_{\mathbf{Q}_2 \succeq \mathbf{0}} \left\{ \frac{1}{2} \log \left| \mathbf{I} + \frac{\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T}{\mathbf{I} + \mathbf{H}_2\mathbf{Q}_1^*\mathbf{H}_2^T} \right| - \frac{1}{2} \log \left| \mathbf{I} + \frac{\mathbf{H}_1\mathbf{Q}_2\mathbf{H}_1^T}{\mathbf{I} + \mathbf{H}_1\mathbf{Q}_1^*\mathbf{H}_1^T} \right| \right\}, \tag{15a}$$

$$\text{s.t.} \quad \mathrm{tr}(\mathbf{Q}_2) \leq P_2 = (1 - \alpha)P. \tag{15b}$$

Since $\mathbf{Q}_1^*$ is given after solving (10), next we show that the problem (15) can be seen as a wiretap channel where users 2 and 1 are the legitimate user and eavesdropper, respectively.

Theorem 2: [38] The above channel can be converted to a standard MIMO wiretap channel with

$$\ddot{\mathbf{H}}_1 \triangleq \mathbf{D}_a^{-\frac{1}{2}}\mathbf{E}_a^T\mathbf{H}_1, \tag{16a}$$

$$\ddot{\mathbf{H}}_2 \triangleq \mathbf{D}_b^{-\frac{1}{2}}\mathbf{E}_b^T\mathbf{H}_2, \tag{16b}$$

in which $\mathbf{D}_a$ and $\mathbf{E}_a$ are the eigenvalues and eigenvectors of $\mathbf{I} + \mathbf{H}_1\mathbf{Q}_1^*\mathbf{H}_1^T$, and $\mathbf{D}_b$ and $\mathbf{E}_b$ are the eigenvalues and eigenvectors of $\mathbf{I} + \mathbf{H}_2\mathbf{Q}_1^*\mathbf{H}_2^T$.

Then, the rate for user 2 can be written as

$$R_{2c}(\alpha_2) = \max_{\mathbf{Q}_2 \succeq \mathbf{0}} \frac{1}{2} \log \frac{|\mathbf{I} + \ddot{\mathbf{H}}_2\mathbf{Q}_2\ddot{\mathbf{H}}_2^T|}{|\mathbf{I} + \ddot{\mathbf{H}}_1\mathbf{Q}_2\ddot{\mathbf{H}}_1^T|}, \tag{17}$$

From (17), it is seen that similar to (10a), (15a) is the rate for a MIMO wiretap channel with channels $\ddot{\mathbf{H}}_2$ for the legitimate user and $\ddot{\mathbf{H}}_1$ for the eavesdropper. This problem now transfers to a MIMO wiretap channel, and we can obtain $\mathbf{Q}_2^*$ using any standard MIMO wiretap solutions.

Step 4: After distributing the power to both users for secrecy messages, we allocate the remaining power $P_0 = \alpha_0 P, \alpha_0 = 1 - \alpha_1 - \alpha_2$ to the common message M_0 for both users. The (5a), (7a), and (8a) becomes

$$R_0(\alpha_0) = \max_{\mathbf{Q}_0 \succeq \mathbf{0}} \min\{R_{0j}\}, j = 1, 2 \tag{18a}$$

$$\text{s.t.} \quad \mathrm{tr}(\mathbf{Q}_0) \leq P_0 = \alpha_0 P, \tag{18b}$$

Since $\mathbf{Q}_1^*$ and $\mathbf{Q}_2^*$ are given, we can show that the above problem becomes MIMO multicasting [18] by applying the same approach as Theorem 1 again into (6). Specifically, let us define the denominator of (6) as

$$\mathbf{K}_j \triangleq \mathbf{I} + \mathbf{H}_j(\mathbf{Q}_1^* + \mathbf{Q}_2^*)\mathbf{H}_j^T \triangleq \mathbf{F}_j\mathbf{G}_j\mathbf{F}_j^T, \tag{19}$$

for $j = 1, 2$, where the second equality is given by eigenvalue decomposition. Then, R_{0j} can be rewritten as

$$\begin{aligned} R_{0j} &= \frac{1}{2} \log \left|\mathbf{I} + \mathbf{K}_j^{-1}\mathbf{H}_j\mathbf{Q}_0\mathbf{H}_j^T\right|, \\ &= \frac{1}{2} \log \left|\mathbf{I} + \mathbf{G}_j^{-\frac{1}{2}}\mathbf{F}_j^T\mathbf{H}_j\mathbf{Q}_0\mathbf{H}_j^T\mathbf{F}_j\mathbf{G}_j^{-\frac{1}{2}}\right|, \\ &= \frac{1}{2} \log |\mathbf{I} + \dddot{\mathbf{H}}_j\mathbf{Q}_0\dddot{\mathbf{H}}_j^T|. \end{aligned} \tag{20}$$

See the proof in Theorem 1. Then we have $\dddot{\mathbf{H}}_j = \mathbf{G}_j^{-\frac{1}{2}}\mathbf{F}_j^T\mathbf{H}_j$, $j = 1, 2$.

The problem (18) with (20) is now identified as the MIMO multicasting which is to maximize the minimum user rate configuration, and the optimal solution $\mathbf{Q}_0^*$ can be achieved by semi-definite programming (SDP), i.e., CVX, however, it may incur a high computational complexity for multiple users and antennas. As we will see in the next subsection, analytical solutions together with a convex tool for different cases are proposed for multicast transmission.

B. The Signaling Design

We solve each sub-problem in this subsection, i.e., design precoding and power allocation for all the secrecy scenarios. Scenario A is composed of two P2P MIMO and one multicasting; Scenario B consists of one wiretap channel, one P2P MIMO, and one multicasting; Scenario C has two wiretap channels and one multicasting.

Scenario A: (*Step 1* $\rightarrow$ *Step 2a* $\rightarrow$ *Step 3a* $\rightarrow$ *Step 4*)

Problem (9) is a P2P MIMO which is convex and has a closed-form solution given in the following Lemma [28].

Lemma 1: [28] For P2P MIMO problem $\max_{\mathbf{Q}\succcurlyeq\mathbf{0}} \log|\mathbf{I}+\mathbf{H}\mathbf{Q}\mathbf{H}^T|$ under a total power constraint, the optimal solution is given by $\mathbf{Q}^*=\mathbf{\Psi}\mathbf{\Gamma}\mathbf{\Psi}^T$. in which $\mathbf{H}=\mathbf{\Phi}\text{diag}(\tau_1,\tau_2,\ldots,\tau_n)\mathbf{\Psi}^T$, $\tau_i\geq 0$, $\forall i$, $\mathbf{\Gamma}=\text{diag}[(\mu-1/\tau_1^2)^+,\ldots,(\mu-1/\tau_n^2)^+]$, μ is the water level.

The solutions of (9) in *Step 2a* and (11) in *Step 3a* are achieved by replacing $\mathbf{H}$ in Lemma 1 by $\mathbf{H}_1$ and $\dot{\mathbf{H}}_2$, respectively, using Theorem 1.

To precode for the common message M_0 in *Step 4*. Define the optimal precoding matrices $\mathbf{Q}_{01}^*$ and $\mathbf{Q}_{02}^*$ for R_{01} and R_{02} in (20), respectively, then we have [1]

- *Case 1:* $R_{01}(\mathbf{Q}_{01}^*)\leq R_{02}(\mathbf{Q}_{01}^*)$, then the optimal multicast covariance matrix of (18) is $\mathbf{Q}_0^*:=\mathbf{Q}_{01}^*$.
- *Case 2:* $R_{01}(\mathbf{Q}_{02}^*)\geq R_{02}(\mathbf{Q}_{02}^*)$, the optimal multicast covariance matrix of (18) is $\mathbf{Q}_0^*:=\mathbf{Q}_{02}^*$.
- *Case 3:* Otherwise, the optimal multicast covariance matrix of (18) can be obtained by a random search.

For *Case 1* or *Case 2*, Lemma 1 [28] is applied. For *Case 3*, optimal $\mathbf{Q}_0^*$ happens when the two convex functions are equal. Then, we can generate $\mathbf{Q}_0$ using the rotation method and search the parameters non-linearly [41].

Finally, DPC rate region $\mathcal{R}_{21}^{\text{DPC}}$ can be reached by exhaustively searching over all power fractions α_1, α_2 and α_0. For each pair of power splitting parameters α_1, α_2, and α_0, we solve precoding matrices $\mathbf{Q}_1^*$, $\mathbf{Q}_2^*$, and $\mathbf{Q}_0^*$ (and thus $R_{1p}(\alpha_1)$, $R_{2p}(\alpha_2)$, and $R_0(\alpha_0)$). Alternatively, $\mathcal{R}_{21}^{\text{DPC}}$ is obtained by encoding the private messages for user 2 and user 1, then the common message for both. We can solve $\mathbf{Q}_2^*$ followed by $\mathbf{Q}_1^*$ and $\mathbf{Q}_0^*$ to obtain $\bar{R}_{1p}(\alpha_1)$, $\bar{R}_{2p}(\alpha_2)$, and $\bar{R}_0(\alpha_0)$, respectively.

Corollary 1: The achievable DPC rate region for secure MIMO-NOMA Scenario A under the total power is the convex hull of all rate triples

$$R_A(P)=\text{conv}\left\{\left(\bigcup_{\alpha_k}\mathcal{R}_{12}^{\text{DPC}}(\alpha_k)\right)\bigcup\left(\bigcup_{\alpha_k}\mathcal{R}_{21}^{\text{DPC}}(\alpha_k)\right)\right\}, \tag{21}$$

$\mathcal{R}_{12}^{\text{DPC}}(\alpha_k)=(R_{1p}^*(\alpha_1),R_{2p}^*(\alpha_2),R_0^*(\alpha_0))$, $k=0,1,2$, and is obtained by encoding the private messages for first user 1 then user 2 followed by the common message for both, whereas $\mathcal{R}_{21}^{\text{DPC}}(\alpha_k)=(\bar{R}_{1p}^*(\alpha_1),\bar{R}_{2p}^*(\alpha_2),\bar{R}_0^*(\alpha_0))$ is obtained in the reverse order of private messages (first user 2 then user 1).

Scenario B: (*Step 1* $\rightarrow$ *Step 2b* $\rightarrow$ *Step 3a* $\rightarrow$ *Step 4*)

Standard MIMO wiretap solutions can be applied to design covariance matrix $\mathbf{Q}_1$ for *Step 2b*. One fast approach is rotation-based linear precoding [31]. In this method, the covariance matrix $\mathbf{Q}_1$ is eigendecomposed into one rotation matrix $\mathbf{V}_1$ and one power allocation matrix $\mathbf{\Lambda}_1$ [31], [41] as

$$\mathbf{Q}_1=\mathbf{V}_1\mathbf{\Lambda}_1\mathbf{V}_1^T. \tag{22}$$

Consequently, the secrecy capacity of user 1 is

$$R_{1c}(\alpha_1)=\max_{\mathbf{Q}_1\succeq\mathbf{0}}\frac{1}{2}\log\frac{|\mathbf{I}+\mathbf{H}_1\mathbf{V}_1\mathbf{\Lambda}_1\mathbf{V}_1^T\mathbf{H}_1^T|}{|\mathbf{I}+\mathbf{H}_2\mathbf{V}_1\mathbf{\Lambda}_1\mathbf{V}_1^T\mathbf{H}_2^T|}, \tag{23a}$$

$$\text{s.t.}\quad\sum_{n=1}^{n_t}\lambda_{1n}\leq P_1=\alpha P, \tag{23b}$$

in which λ_{1n}, $n=\{1,\ldots,n_t\}$, is a diagonal element of matrix $\mathbf{\Lambda}_1=\text{diag}(\lambda_{11},\ldots,\lambda_{1n_t})$. The rotation matrix $\mathbf{V}_1$ can be obtained by

$$\mathbf{V}_1=\prod_{p=1}^{n_t-1}\prod_{q=p+1}^{n_t}\mathbf{V}_{pq}, \tag{24}$$

in which the basic rotation matrix $\mathbf{V}_{pq}$ is a Givens matrix which is an identity matrix except that its elements in the pth row and qth column, i.e., v_{pp}, v_{pq}, v_{qp}, and v_{qq} are replaced by

$$\begin{bmatrix}v_{pp} & v_{pq}\\ v_{qp} & v_{qq}\end{bmatrix}=\begin{bmatrix}\cos\theta_{1pq} & -\sin\theta_{1pq}\\ \sin\theta_{1pq} & \cos\theta_{1pq}\end{bmatrix}, \tag{25}$$

in which θ_{1pq} is rotation angle corresponding to the rotation matrix $\mathbf{V}_{pq}$. Then, we will optimize the parameterized problem by applying numerical approaches such as Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [43] to obtain the solution $\mathbf{Q}_1^*$ (thus $R_{1c}^*(\alpha_1)$). To obtain $\mathbf{Q}_2^*$ and $R_{2p}^*(\alpha_2)$ in *Step 3a*, $\mathbf{Q}_1^*$ above is applied in Theorem 1, and we solve the modified P2P MIMO problem using Lemma 1. The precoding approach for *Step 4* is the same as Scenario A. The achievable secrecy rate for Scenario B is given by the following corollary.

Corollary 2: The achievable rate region for secure MIMO-NOMA Scenario B under the total power is the convex hull of all rate triples

$$R_B(P)=\bigcup_{\alpha_k}(R_{1c}^*(\alpha_1),R_{2p}^*(\alpha_2),R_0^*(\alpha_0)). \tag{26}$$

Scenario C: (*Step 1* $\rightarrow$ *Step 2b* $\rightarrow$ *Step 3b* $\rightarrow$ *Step 4*)

In Scenario C, the steps are the same as Scenario B except for *Step 3b* which can be seen as a wiretap channel instead of P2P MIMO. Then, we apply Theorem 2 and solve (17) instead. Similar to the precoding in *Step 2b* of Scenario B, the covariance matrix $\mathbf{Q}_2$ can be written by rotation method as $\mathbf{Q}_2=\mathbf{V}_2\mathbf{\Lambda}_2\mathbf{V}_2^T$, where the rotation matrix $\mathbf{V}_2$ is defined similarly to $\mathbf{V}_1$ in (24) with rotation angles are θ_{2pq}. Therefore, the optimization problem for $R_{2c}(\alpha_2)$ becomes

$$R_{2c}(\alpha_2)=\max_{\mathbf{Q}_2\succeq\mathbf{0}}\frac{1}{2}\log\frac{|\mathbf{I}+\ddot{\mathbf{H}}_2\mathbf{V}_2\mathbf{\Lambda}_2\mathbf{V}_2^T\ddot{\mathbf{H}}_2^T|}{|\mathbf{I}+\ddot{\mathbf{H}}_1\mathbf{V}_2\mathbf{\Lambda}_2\mathbf{V}_2^T\ddot{\mathbf{H}}_1^T|}, \tag{27a}$$

$$\text{s.t.}\quad\sum_{n=1}^{n_t}\lambda_{2n}\leq P_2=(1-\alpha)P, \tag{27b}$$

in which $\mathbf{\Lambda}_2=\text{diag}(\lambda_{21},\ldots,\lambda_{2n_t})$. This problem is again similar to (23).

In the power splitting scheme, we solve $\mathbf{Q}_1^*$, $\mathbf{Q}_2^*$, and $\mathbf{Q}_0^*$ to obtain $R_{1c}^*(\alpha_1)$, $R_{2c}^*(\alpha_2)$, and $R_0^*(\alpha_0)$ with respect to power splitting parameters pair $(\alpha_1,\alpha_2,\alpha_0)$. Alternatively, we can first solve for $\mathbf{Q}_2^*$ followed by $\mathbf{Q}_1^*$ last $\mathbf{Q}_0^*$ (i.e., first $\bar{R}_{1c}^*(\alpha_1)$, $\bar{R}_{2c}^*(\alpha_2)$, then $\bar{R}_0^*(\alpha_0)$). In general, changing the order

Algorithm 1: Power Splitting for all Three Scenarios.

1: inputs: secrecy scenario $\mathrm{L} \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$, and ϵ_1;
2: **for** $\alpha_1 = 0 : \epsilon_1 : 1$ **do**
3:   **for** $\alpha_2 = 0 : \epsilon_1 : 1 - \alpha_1$ **do**
4:     $\alpha_0 = 1 - \alpha_1 - \alpha_2$;
5:     **switch** L
6:     **case** A**:**
7:       Obtain $\mathbf{Q}_1^*$ using Lemma 1 in problem (9);
8:       Compute R_{1p} in (9);
9:     **case** B or C**:**
10:       Obtain $\mathbf{Q}_1^*$ by solving (23) using BFGS;
11:       Compute R_{1c} in (10);
12:     **end switch**
13:     **switch** L
14:     **case** A or B**:**
15:       Obtain $\mathbf{Q}_2^*$ using Theorem 1, the $\mathbf{Q}_1^*$ in Line 7 or Line 10, and Lemma 1 in problem (11);
16:       Compute R_{2p} in (11);
17:     **case** C**:**
18:       Obtain $\mathbf{Q}_2^*$ using Theorem 2, the $\mathbf{Q}_1^*$ in Line 10, and BFGS by solving (27);
19:       Compute R_{2c} in (15);
20:     **end switch**
21:     Compute R_0 as described in *Step 4*;
22:   **end for**
23: **end for**
24: **if** $\mathrm{L} = \mathrm{A}$ or $\mathrm{L} = \mathrm{C}$ **then**
25:   swap all subscripts of 1 and 2 in (5) or (8);
26:   repeat **switch** and obtain $\mathcal{R}_{21}^{\mathrm{DPC}}(\alpha_k)$ or $\mathcal{R}_{21}(\alpha_k)$ in Corollary 1 and Corollary 3;
27: **end if**
28: outputs: $R_{\mathrm{L}}(P)$.

of optimization will result in a different rate region. The *convex hull* of the two solutions with different orders enlarges the achievable rate region. The achievable secrecy rate for Scenario C is given by Corollary 3.

Corollary 3: The achievable S-DPC rate region for the secure MIMO-NOMA Scenario C under the total power is the convex hull of all rate triples

$$R_C(P) = \mathrm{conv}\left\{ \left(\bigcup_{\alpha_k} \mathcal{R}_{12}(\alpha_k) \right) \bigcup \left(\bigcup_{\alpha_k} \mathcal{R}_{21}(\alpha_k) \right) \right\}, \tag{28}$$

in which $\mathcal{R}_{12}(\alpha_k) = (R_{1c}^*(\alpha_1), R_{2c}^*(\alpha_2), R_0^*(\alpha_0))$, $k = 0, 1, 2$, is obtained by encoding the confidential messages for user 1 first, then user 2, and lastly the common message for both, whereas $\mathcal{R}_{21}(\alpha_k) = (\bar{R}_{1c}^*(\alpha), \bar{R}_{2c}^*(\alpha), \bar{R}_0^*(\alpha_0))$ is obtained in the reverse order of confidential messages (first user 2 then user 1).

Algorithm 1 summarizes the power splitting method for all scenarios. ϵ_1 is the searching step for the power allocation factor. If $\alpha_1 = \alpha_2 = 0$ and $\alpha_0 \neq 0$, then the system reduces to multicasting transmission. If no power is allocated to the common message, it is the private transmission cases in the next Section IV. If only the power of one of the secrecy messages is zero ($\alpha_k = 0$, $k = 1$ or 2), the problem is the integrated service with confidential and common messages [41].

The precoding order for secrecy messages at different scenarios is not the same. Corollary 1 and Corollary 3 require an exchange of subscripts. For Scenario A, this is because the encoding order affects the achievable rate region. For Scenario C, although encoding order is irrelevant to the achievable rate in S-DPC, the order of optimization (solve the covariance matrix) will affect the solution [38]. This is because the power splitting method splits the power among the messages and solves them one by one. This simplifies the problem but is sub-optimal in general. Then, changing the precoding order may enlarge the achievable rate region. For scenario B, as proved in [23, Remark 4], it is always better to cancel the private massage M_{2p} at user 1 and treat the confidential message M_{1c} at user 2 as noise [23, Remark 4]. Thus, there is no need to exchange the precoding order.

Remark 1 (Complexity): For Scenario A, *Step 2a*, *Step 3a*, Case 1, and Case 2 in *Step 4* are analytical, which only requires the computation of matrix multiplications and matrix inverse. The computation of matrix multiplications and matrix inverse has the complexity of $\mathcal{O}(m^3)$ in which $m = \max(n_t, n_1, n_2)$. Case 3 in *Step 4* uses `fmincon` which is achieved mainly by BFGS. The BFGS algorithm yields the complexity $\mathcal{O}(n^2)$ [43], and the input variable $n = \frac{(n_t+1)n_t}{2}$ is rotation parameters [44]. Thus, the complexity of Scenario A is $\mathcal{O}(\frac{m^3+n_t^4}{\epsilon_1^2})$. For Scenario B and Scenario C, the complexity of solving wiretap channels in *Step 2b* and *Step 3b* has $\mathcal{O}(m^3 + n_t^4)$. Ignore the coefficient, the overall complexity of Algorithm 1 is $\mathcal{O}(\frac{m^3+n_t^4}{\epsilon_1^2})$. The achievable DPC/S-DPC rate region in Scenario A [20], Scenario B [23], and Scenario C [15], [24] are found by using an exhaustive search over a set of positive semidefinite matrices, which have exponential complexity in terms of m. The three-dimensional space search in DPC or S-DPC has to be "exhaustive" but the search over the power allocation factors is linear.

IV. Weighted Sum Rate Formulation for Secrecy

We consider the subcases of the three scenarios without a common message ($M_0 = \emptyset$) in this section. A WSR maximization based on BSMM [37], [42] is generalized to all scenarios. The WSR maximization under a total power constraint is formulated as

$$\varphi(P) = \max_{\mathbf{Q}_j \succcurlyeq \mathbf{0}} \sum_j w_j R_j, \quad j = 1, 2$$
$$\text{s.t.} \quad \mathrm{tr}(\mathbf{Q}_j) \leq P, \tag{29}$$

where $R_j := R_{jp}$ in Scenario A, $R_1 := R_{1c}$ and $R_2 := R_{2p}$ in Scenario B, and $R_j := R_{jc}$ in Scenario C. $w_j \geq 0$ is a weight. The Lagrangian of the problem (29) is

$$L(\mathbf{Q}_1, \mathbf{Q}_2, \lambda) = w_1 R_1 + w_2 R_2 - \lambda(\mathrm{tr}(\mathbf{Q}_1 + \mathbf{Q}_2) - P), \tag{30}$$

where λ is the Lagrange multiplier related to the total power constraint. The dual function is a maximization of the Lagrangian

$$g(\lambda) = \max_{\mathbf{Q}_k \succcurlyeq \mathbf{0}} L(\mathbf{Q}_1, \mathbf{Q}_2, \lambda), \tag{31}$$

and the dual problem is given by $\min_{\lambda \geq 0} g(\lambda)$.

Lemma 2: The problem in (29) has zero duality gap and the KKT conditions are necessary for the optimal solution.

Proof: The duality gap is zero in Scenario A because the problem can be transferred to a convex problem satisfying Slater's condition [45]. Scenario B has zero duality gap, see the details in Appendix A in [2]. Scenario C has been discussed in [37, Theorem 1] which satisfies Lemma 2. □

Since the problem in (31) is a nonconvex problem in any secrecy transmission, the BSMM [37], [42] can be considered which alternatively updates covariance matrix by maximizing a set of strictly convex local approximations. Specifically, Scenario C has been studied in [37]. We discuss Scenario A and Scenario B in this paper.

A. Scenario A

It is worth noting that the MAC-BC duality [32] is applied to the WSR maximization in [34] where the WSR on the MAC rate region is transformed to an equivalent WSR on the BC rate region by an iterative algorithm. Then, the WSR can be solved using convex optimization. Once the optimum uplink covariance matrices are determined by any standard convex optimization tool, the equivalent downlink covariance matrices can be obtained through the duality transformation [32]. The optimization in MAC requires a descent algorithm over a line search with a tolerance. It also mentions that the DPC rate region is difficult to compute without employing duality [32]. Yet in this paper, we provide an alternative solution without applying the MAC-BC duality. We form a WSR for the DPC rate region directly and solve the maximization by using BSMM.

We can apply BSMM which updates covariance matrices by successively optimizing the lower bound of local approximation of $f(\mathbf{Q}_1, \mathbf{Q}_2) = L(\mathbf{Q}_1, \mathbf{Q}_2, \lambda)$ [37], [42]. Rewrite $f(\mathbf{Q}_1, \mathbf{Q}_2)$ into the summation of one convex and one concave functions

$$f(\mathbf{Q}_1, \mathbf{Q}_2) = f_1(\mathbf{Q}_1) + f_2(\mathbf{Q}_1, \mathbf{Q}_2), \tag{32}$$

in which

$$f_1(\mathbf{Q}_1) = \frac{w_1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{Q}_1 \mathbf{H}_1^T| - \lambda \mathrm{tr}(\mathbf{Q}_1) \tag{33a}$$

$$\begin{aligned} f_2(\mathbf{Q}_1, \mathbf{Q}_2) = \frac{w_2}{2} \log |\mathbf{I} + (\mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1 \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{Q}_2 \mathbf{H}_2^T| \\ - \lambda(\mathrm{tr}(\mathbf{Q}_2) - P). \end{aligned} \tag{33b}$$

$f_1(\mathbf{Q}_1)$ is a concave function of $\mathbf{Q}_1$, $f_2(\mathbf{Q}_1, \mathbf{Q}_2)$ is convex over $\mathbf{Q}_1$ by fixing $\mathbf{Q}_2$. After the decomposition, we can alternatively optimize $\mathbf{Q}_1$ and $\mathbf{Q}_2$ to find a lower bound for the weighted sum rate. For the ith iteration, the function for $f_2(\mathbf{Q}_1, \mathbf{Q}_2^{(i-1)})$ is lower-bounded by its first-order Taylor approximation [45]

$$\begin{aligned} f_2(\mathbf{Q}_1, \mathbf{Q}_2^{(i-1)}) \geq f_2(\mathbf{Q}_1^{(i-1)}, \mathbf{Q}_2^{(i-1)}) \\ -\mathrm{tr}[\mathbf{A}(\mathbf{Q}_1 - \mathbf{Q}_1^{(i-1)})] \end{aligned} \tag{34}$$

in which the power price matrix $\mathbf{A}$ is a negative partial derivative with respect to $\mathbf{Q}_1$

$$\begin{aligned} \mathbf{A} &= -\nabla_{\mathbf{Q}_1} f_2(\mathbf{Q}_1^{(i-1)}, \mathbf{Q}_2^{(i-1)}) \\ &= -\frac{w_2}{\ln 2} \mathbf{H}_2^T (\mathbf{I} + \mathbf{H}_2(\mathbf{Q}_1^{(i-1)} + \mathbf{Q}_2^{(i-1)}) \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \\ &\quad + \frac{w_2}{\ln 2} \mathbf{H}_2^T (\mathbf{I} + \mathbf{H}_2 (\mathbf{Q}_1^{(i-1)}) \mathbf{H}_2^T)^{-1} \mathbf{H}_2. \end{aligned} \tag{35}$$

Then the problem is lower bounded as

$$\begin{aligned} f(\mathbf{Q}_1, \mathbf{Q}_2^{(i-1)}) \geq f_1(\mathbf{Q}_1) + f_2(\mathbf{Q}_1^{(i-1)}, \mathbf{Q}_2^{(i-1)}) \\ - \mathrm{tr}[\mathbf{A}(\mathbf{Q}_1 - \mathbf{Q}_1^{(i-1)})]. \end{aligned} \tag{36}$$

Then, we optimize the right-hand side of (36) by omitting the constant terms, which is equivalent as

$$\mathbf{Q}_1^{(i)} = \arg\max_{\mathbf{Q}_1} \frac{w_1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{Q}_1 \mathbf{H}_1^T| - \mathrm{tr}[(\lambda \mathbf{I} + \mathbf{A})\mathbf{Q}_1]. \tag{37}$$

Next, we optimize $f(\mathbf{Q}_1^{(i)}, \mathbf{Q}_2)$ by fixing $\mathbf{Q}_1^{(i)}$, which is equivalent as

$$\begin{aligned} \mathbf{Q}_2^{(i)} = \arg\max_{\mathbf{Q}_2} \frac{w_2}{2} \log |\mathbf{I} + (\mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1^{(i)} \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{Q}_2 \mathbf{H}_2^T| \\ - \lambda \mathrm{tr}(\mathbf{Q}_2). \end{aligned} \tag{38}$$

The optimal solution for (37) and (38) can be achieved by the following lemma [37].

Lemma 3: [37] For some $\mathbf{S} \succ \mathbf{0}$, the optimal solution of the problem

$$\max_{\mathbf{Q} \succcurlyeq \mathbf{0}} w \log |\mathbf{I} + \mathbf{R}^{-1} \mathbf{H} \mathbf{Q} \mathbf{H}^T| - \mathrm{tr}(\mathbf{S}\mathbf{Q}) \tag{39}$$

is given by

$$\mathbf{Q}^* = \mathbf{S}^{-1/2} \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \mathbf{S}^{-1/2}. \tag{40}$$

To use Lemma 3, we set $w = \frac{w_1}{2}$, $\mathbf{S} = \lambda \mathbf{I} + \mathbf{A}$, and $\mathbf{R} = \mathbf{I}$ for (37); and $w = \frac{w_2}{2}$, $\mathbf{S} = \lambda \mathbf{I}$, and $\mathbf{R} = \mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1^{(i)} \mathbf{H}_2^T$ for (38). $\mathbf{V}$, $\mathbf{U}$, and $\mathbf{\Lambda}$ are obtained by eigenvalue decomposition of $\mathbf{R}^{-1/2} \mathbf{H} \mathbf{S}^{-1/2} = \mathbf{U} \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m) \mathbf{V}^T$, $\sigma_i \geq 0$, $\forall i$, $\mathbf{\Lambda} = \mathrm{diag}[(w - 1/\sigma_1^2)^+, \ldots, (w - 1/\sigma_m^2)^+]$, and $(x)^+ = \max(x, 0)$.

B. Scenario B

In Scenario B, what makes it different from Scenario A is the formulation of convex and concave functions, which can be written as

$$f_1(\mathbf{Q}_1) = \frac{w_1}{2} \log |\mathbf{I} + \mathbf{H}_1 \mathbf{Q}_1 \mathbf{H}_1^T| - \lambda \mathrm{tr}(\mathbf{Q}_1) \tag{41a}$$

$$\begin{aligned} f_2(\mathbf{Q}_1, \mathbf{Q}_2) = &-\frac{w_1}{2} \log |\mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1 \mathbf{H}_2^T| \\ &+ \frac{w_2}{2} \log |\mathbf{I} + (\mathbf{I} + \mathbf{H}_2 \mathbf{Q}_1 \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{Q}_2 \mathbf{H}_2^T| \\ &- \lambda(\mathrm{tr}(\mathbf{Q}_2) - P). \end{aligned} \tag{41b}$$

$f_1(\mathbf{Q}_1)$ is also a concave function of $\mathbf{Q}_1$, $f_2(\mathbf{Q}_1, \mathbf{Q}_2)$ is convex by fixing $\mathbf{Q}_2$ because the second term in (41b) is convex over $\mathbf{Q}_1$. For the ith iteration, the function for $f_2(\mathbf{Q}_1, \mathbf{Q}_2^{(i-1)})$ is lower-bounded by its first-order Taylor approximation as the expression in (34), in which the power price matrix

$$\mathbf{A} = -\nabla_{\mathbf{Q}_1} f_2(\mathbf{Q}_1^{(i-1)}, \mathbf{Q}_2^{(i-1)})$$

Algorithm 2: WSR Maximization for all Three Scenarios Without a Common Message.

inputs: $\lambda^{\max}$, $\lambda^{\min}$, ϵ_2, ϵ_3, secrecy scenario $\mathrm{L} \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$;
while $\lambda^{\max} - \lambda^{\min} > \epsilon_2$ **do**
  $\lambda := (\lambda^{\max} + \lambda^{\min})/2$;
  $\mathbf{Q}_1^{(0)} := \mathbf{Q}_2^{(0)} := \frac{P}{2n_t}\mathbf{I}$;
  $R^{(0)} := 0$;
  $i = 0$;
  while 1 **do**
    $i = i + 1$;
    switch L
    case A:
      Solve $\mathbf{Q}_1^{(i)}$ and $\mathbf{Q}_2^{(i)}$ in (37)-(38) using Lemma 3;
      Compute R_1 and R_2 in (5);
    case B:
      Solve $\mathbf{Q}_1^{(i)}$ and $\mathbf{Q}_2^{(i)}$ in (43)-(44) using Lemma 3;
      Compute R_1 and R_2 in (7);
    case C:
      Solve $\mathbf{Q}_1^{(i)}$ and $\mathbf{Q}_2^{(i)}$ in [37, Algorithm 1, lines 5- 13];
      Obtain R_1 and R_2 in (8);
    end switch
    $R^{(i)} := w_1 R_1 + w_2 R_2$
    if $\mathrm{abs}(R^{(i)} - R^{(i-1)}) < \epsilon_3$ **then**
      break;
    end if
    if $\mathrm{tr}(\mathbf{Q}_1^{(i)} + \mathbf{Q}_2^{(i)}) < P$ **then**
      $\lambda^{\max} := \lambda$;
    else
      $\lambda^{\min} := \lambda$;
    end if
  end while
end while
outputs: $\lambda^* := \lambda$, $R_k^* := R_k$, and $\mathbf{Q}_k^* = \mathbf{Q}_k^{(i)}$, $k \in \{1, 2\}$.

$$= \frac{w_1 + w_2}{2\ln 2}\mathbf{H}_2^T(\mathbf{I} + \mathbf{H}_2\mathbf{Q}_1^{(i-1)}\mathbf{H}_2^T)^{-1}\mathbf{H}_2 - \frac{w_2}{2\ln 2}\mathbf{H}_2^T(\mathbf{I} + \mathbf{H}_2(\mathbf{Q}_1^{(i-1)} + \mathbf{Q}_2^{(i-1)})\mathbf{H}_2^T)^{-1}\mathbf{H}_2. \tag{42}$$

Finally, we optimize the right-hand side of (36) with the power price matrix in (42), which is equivalent as

$$\mathbf{Q}_1^{(i)} = \arg\max_{\mathbf{Q}_1} \frac{w_1}{2}\log|\mathbf{I} + \mathbf{H}_1\mathbf{Q}_1\mathbf{H}_1^T| - \mathrm{tr}[(\lambda\mathbf{I} - \mathbf{A})\mathbf{Q}_1]. \tag{43}$$

Next, we optimize $L(\mathbf{Q}_1^{(i)}, \mathbf{Q}_2)$ by fixing $\mathbf{Q}_1^{(i)}$, which is equivalent as

$$\mathbf{Q}_2^{(i)} = \arg\max_{\mathbf{Q}_2} \frac{w_2}{2}\log|\mathbf{I} + (\mathbf{I} + \mathbf{H}_2\mathbf{Q}_1^{(i)}\mathbf{H}_2^T)^{-1}\mathbf{H}_2\mathbf{Q}_2\mathbf{H}_2^T| - \lambda\mathrm{tr}(\mathbf{Q}_2). \tag{44}$$

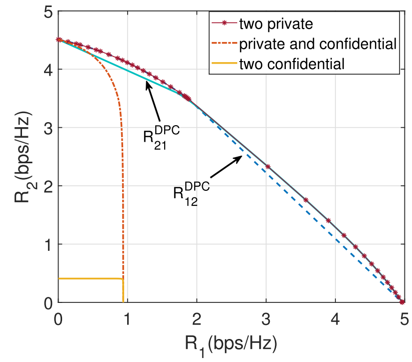


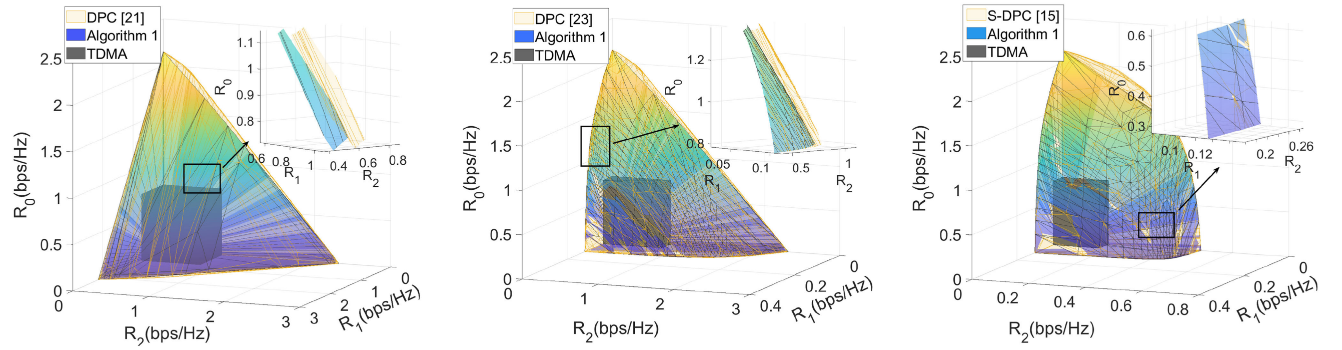
Fig. 3. Capacity regions of three scenarios under an average total power constraint without common message over the channel $H_1 = [0.3\ 2.5; 2.2\ 1.8]$ and $H_2 = [1.3\ 1.2; 1.5\ 3.9]$, and $P = 12$.

The WSR maximization for all scenarios without a common message is summarized in Algorithm 2. ϵ_2 and ϵ_3 are the bisection search accuracy and convergence tolerance of BSMM, respectively. If $w_1 = 0$ and $w_2 = 1$, the problem reduces to a P2P MIMO with an analytical solution. Algorithm 2 becomes a WF regime. If $w_1 = 1$ and $w_2 = 0$, then the problem reduces to a MIMO wiretap channel. Then, Algorithm 2 is nothing but AOWF [30]. The WSR maximization is hard to be extended directly to the general cases of the three scenarios, because the max-min problem of multicasting is not derivable in BSMM although the multicasting problem owns convexity.

Encoding order in different scenarios is distinguished. In Scenario A, the weight determines the optimal encoding order. For example, if $w_1 > w_2$, the optimal encoding order is to encode user 1 first and then user 2. In Scenario B, the entire capacity region uses DPC to cancel the signal of the private message M_{2p} intended for user 2 at user 1 only. The other variant which treats the private message M_{2p} of user 2 as interference for user 1 is unnecessary [23, Remark 4]. In Scenario C, the S-DPC owns the invariant property that the achievable rate region is irrelevant to the encoding order [15].

The three scenarios without common messages differentiate the security requirements. For comparison, we show an example in Fig. 3 with the same channel settings as [14], [23]. First, when the secrecy message of user 2 is empty, i.e., $M_{2p} = \emptyset$ in Scenario B and $M_{2c} = \emptyset$ in Scenario C, the maximal achieving rates for user 1 in the two cases are the same, and the two problems drop to the Gaussian wiretap channel. Second, when the secrecy message of user 1 is empty, i.e., $M_{1p} = \emptyset$ in Scenario A and $M_{1c} = \emptyset$ in Scenario B, the achieving rates for user 2 in the two cases become the same P2P MIMO problem. Third, imposing a secrecy constraint on two users in Scenario C strictly shrinks the capacity region compared with Scenario A.

Remark 2 (Complexity): The number of iterations of the BSMM is $\mathcal{O}(1/\epsilon_3)$, and the bisection search requires $\mathcal{O}(\log(1/\epsilon_2)$. The WSR of Algorithm 2 has the complexity of $\mathcal{O}(\frac{m^3}{\sigma\epsilon_3}\log(1/\epsilon_2))$ with a search step σ over the weight [37], [38]. On the other hand, the computation complexity of Algorithm 1 without common messages is $\mathcal{O}(\frac{m^3+n_t^4}{\epsilon_1})$ with only one layer of search loop over α_1.



(a) Scenario A (b) Scenario B (c) Scenario C

Fig. 4. Secrecy rate regions of MIMO-NOMA with different scenarios of security ($n_t = n_1 = n_2 = 2$, and $P = 10$ with the same channels shown in (45)). The yellow curved mesh is the secrecy capacity region, the colorful surface denotes the achievable rate region realized by Algorithm 1, and the TDMA (gray cube) is achieved via three orthogonal time slots.

V. Numerical Results

In this section, we perform numerical results to illustrate the achievable secrecy rate region of the three scenarios and then verify Algorithm 1 and Algorithm 2.

A. Secrecy Rate Regions for Three Scenarios

We verify the transmission rates and consider the same channels for all three scenarios, and the channels for user 1 and user 2 chosen to be

$$\mathbf{H}_1 = \begin{bmatrix} 0.3861 \; 0.6355 \\ 0.9995 \; 0.6259 \end{bmatrix}, \; \mathbf{H}_2 = \begin{bmatrix} 0.4977 \; 0.9658 \\ 0.9245 \; 0.6116 \end{bmatrix}, \tag{45}$$

where the channel coefficients are generated randomly according to the standard Gaussian distribution, and the total power is 10. The search steps for α_1 in Algorithm 1 is 0.05. Fig. 4 depicts the secrecy rate regions of the three scenarios. The PS scheme is compared with TDMA based scheme which is realized by transmitting messages in three orthogonal time slots with equal length. Also, the upper bounds are achieved by DPC [20], [21] for Scenario A, capacity rate regions [23] and [15] for Scenario B and C, respectively, which are realized by exhaustive search over all possible covariance matrices. It is shown that the proposed precoding and power allocation method significantly outperforms the TDMA strategy, and it is close to that of the capacity rate regions. The projection of the secrecy capacity region onto the (R_1, R_2) or (R_0, R_j), $j = 1, 2$, plane is the capacity region with two secrecy messages or only one secrecy message, which is going to appear in the next subsections.

It is worth mentioning that in Scenario A, given a set of power allocation parameters, we can analytically obtain the rate triples, i.e., SVD and WF in *Step 2a* and *Step 3a*. The complexity of the algorithm for finding one point on the region only comes from matrix operations, and no search is needed. In [11, Section III] where each user is equipped with one antenna, the rate maximization optimization is transferred to the power minimization problem, and thus a linear semi-definite convex optimization is obtained, but it needs a binomial search of one parameter and then apply one numerical method using standard SDP methods, e.g., CVX [46].

B. Secrecy Rate Regions Without Common Messages

Consider the MIMO-NOMA without a common message. The achievable rate region is realized by Algorithm 1 with $M_0 = \emptyset$ and $\alpha_0 = 0$, and Algorithm 2. The capacity regions are achieved by the parameters including the search step 0.01, the total power $P = 2, 4, 10$, respectively, and the channels for all three scenarios are

$$\mathbf{H}_1 = \begin{bmatrix} 0.1560 & -0.6372 & -0.4055 \\ -1.1450 & -0.1417 & 0.0708 \end{bmatrix},$$

$$\mathbf{H}_2 = \begin{bmatrix} -1.5032 \; 0.5503 \; -0.0334 \end{bmatrix}. \tag{46}$$

Figure 5 compares the rate regions of the proposed power splitting scheme with the capacity region achieved by DPC for Scenario A [6], [33] generated using the iterative algorithm with MAC-BC duality presented [34], and Scenario B [23], respectively, and S-DPC [14] for Scenario C. In Scenario C, our algorithms are compared with the GSVD [36] and BSMM [37]. Both proposed methods can reach the secrecy capacity region and outperform OMA. The PS method in Algorithm 1 is faster and more general for all scenarios, while the WSR method in Algorithm 2 is specific to the case without multicasting.

To illustrate the effectiveness of the algorithms in Scenario A compared with GSVD in [35], we consider another case when the number of receivers' antennas is limited to be the same, i.e., $n_1 = n_2$. The channels are

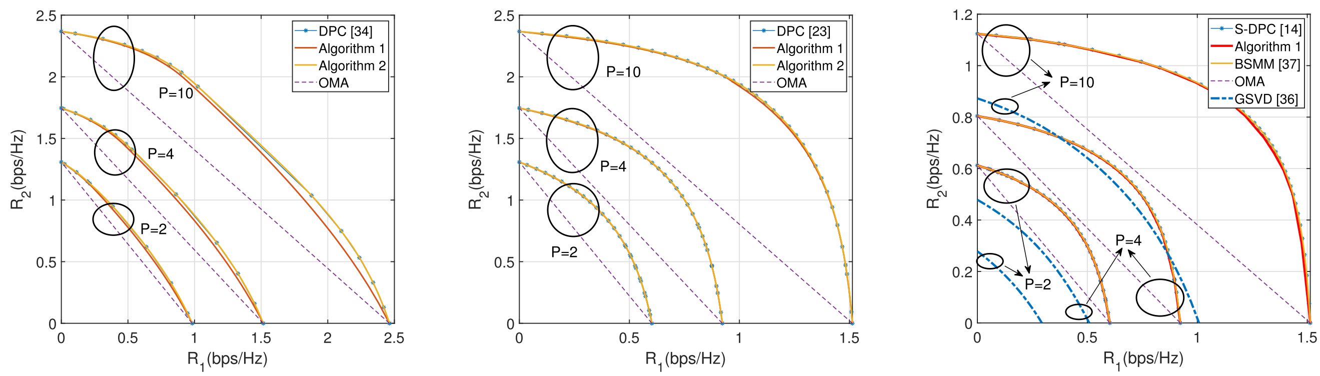
$$\mathbf{H}_1 = \begin{bmatrix} -1.3784 & 0.2593 & -0.2040 \\ -1.0689 & -2.4811 & -1.2978 \end{bmatrix},$$

$$\mathbf{H}_2 = \begin{bmatrix} -0.3403 & 0.1358 & -1.9706 \\ -2.2982 & -1.8135 & 0.2904 \end{bmatrix}, \tag{47}$$

and $P = 10$. From Fig. 6, the proposed algorithms can achieve a larger rate region than GSVD [35] and OMA. In addition, we provide one case with the same setting as in [32, Fig. 3] to show the effectiveness of our algorithms. The channels are:

$$\mathbf{h}_1 = \begin{bmatrix} 1 \; 0.4 \end{bmatrix}, \; \mathbf{h}_2 = \begin{bmatrix} 0.4 \; 1 \end{bmatrix}, \tag{48}$$

and $P = 10$. The results are shown in Fig. 7. The iteration tolerance t in [34] is set as 10^{-3}, and a bisection search is applied to find the optimal t. We set our iteration accuracy ϵ_2 and convergence tolerance ϵ_3 in Algorithm 2 as 10^{-3}. The



(a) Scenario A (b) Scenario B (c) Scenario C

Fig. 5. Secrecy rate regions of MIMO-NOMA without multicasting services with different security requirements ($n_t = 3$, $n_1 = 2$, $n_2 = 1$ and $P = 2, 4, 10$ with the same channel setting in (46)). The blue dot line denotes the achievable or secrecy capacity region realized by DPC or S-DPC, the red line and yellow line are achieved by Algorithm 1 and Algorithm 2, respectively. The dash purple line is OMA reached by the time-sharing between the two extreme points [38].

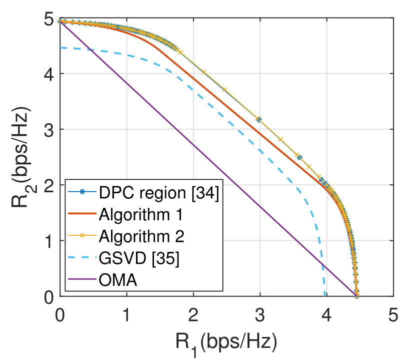


Fig. 6. Comparison of the rate regions of Scenario A, DPC [34], GSVD [35], the proposed schemes, and OMA for $P = 10$, $n_t = 3, n_1 = n_2 = 2$, and channels are given in (47).

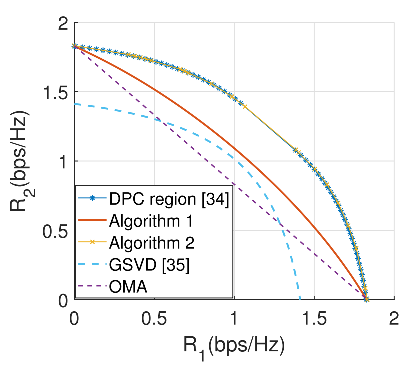


Fig. 7. Comparison of the rate regions of Scenario A, DPC [34], GSVD [35], the proposed schemes, and OMA for $P = 10$, and $n_t = 2, n_1 = n_2 = 1$, and the channels are given in (48).

complexity is the same because both methods require finding the covariance matrices iteratively. The tolerance in [34] and the Lagrange multiplier in Algorithm 2 are both optimized through bisection search. Algorithm 1 is very fast without any search for one power allocation factor but is sub-optimal.

C. Multicast Message and One Confidential Message

If we set $\alpha_2 = 0$ in Scenario C, then the general problem is reduced to the integrated services with one confidential and

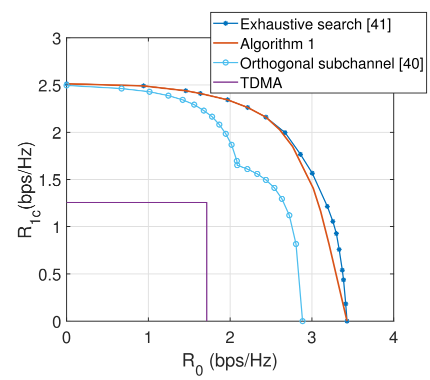


Fig. 8. Comparison of the achievable rate regions of rotation-based exhaustive search [41], GSVD [40], the proposed scheme, and TDMA for $P = 15$, $n_t = 3, n_1 = 4, n_2 = 3$, and channels are given in (49).

one common message[2], i.e., (R_0, R_{1c}). As shown in Fig. 8, the proposed method substantially outperforms the GSVD-based orthogonal subchannel precoding method in [40], in which the turning point is a switch of subchannel selection schemes. Compared with the GSVD, Algorithm 1 makes better use of the channel without decomposing the channel into many orthogonal subchannels. Also, our method is very close to the secrecy capacity obtained by rotation-based random exhaustive search [41]. In this simulation, search step for α_1 is 0.05, $P = 15$, and channels are

$$\mathbf{H}_1 = \begin{bmatrix} 0.0653 & 0.0185 & 1.0397 \\ -0.1762 & -1.5297 & 0.1460 \\ 0.9822 & -1.9882 & -0.1263 \\ 0.9421 & -0.1771 & 0.3746 \end{bmatrix},$$

$$\mathbf{H}_2 = \begin{bmatrix} -0.0248 & 1.3016 & 0.4677 \\ 0.0523 & -0.1297 & 0.4269 \\ 0.6795 & -1.1725 & -0.8358 \end{bmatrix}. \tag{49}$$

We notice that GSVD has been applied to many subcases. Examples are two private messages in Scenario A [35], two confidential messages in Scenario C [36], and one confidential

[2]One can also set $\alpha_1 = 0$ and change the order of channels for (R_0, R_{2c}) which finally will resort to the same results due to duality.

TABLE II
COMPARISON AMONG DIFFERENT PRECODING SCHEMES FOR THE MIMO-NOMA WITH DIFFERENT COMMUNICATION SCENARIOS

	DPC/S-DPC	**WSR with BSMM** (proposed for Scenario A, B without common message)	**PS** (proposed for all scenarios)	**GSVD**	**OMA**
Performance	optimal	suboptimal (but close to optimal)	suboptimal (but close to optimal)	suboptimal	highly suboptimal
Speed	generally slow	acceptable for a small m ($m = \max(n_t, n_1, n_2)$)	fast for a small n_t	fast	very fast
Complexity	generally high	acceptable	acceptable	low	low
Generality	✓	not easy to generalize for common message	✓	✓	✓

message and one common message [47]. Thus, it also has the potential to become an efficient and general tool for all scenarios. But, it should be noted that the performance of GSVD is affected by the number of antennas at the transmitter and users [35], [44]. Algorithm 2 outperforms GSVD and sometimes Algorithm 1, but it is not easy to extend it to common messages. Algorithm 1 balances the two methods. We summarize the benefits and properties of the precoding schemes in Table II.

Three signaling design families in the MIMO-NOMA with different secrecy requirements are:

- GSVD is the fastest general tool but has poor performance in some antenna settings.
- PS (Algorithm 1) is a general but suboptimal tool. It balances time and performance.
- WSR (Algorithm 2) is locally optimal with KKT as the optimal necessary conditions. It has relatively high time complexity.

VI. CONCLUSION

We have investigated a two-user MIMO-NOMA network with different security requirements. Specifically, three scenarios are differentiated according to the required services: multicast, private, and/or confidential services. A PS scheme has been proposed which decomposes the MIMO-BC into the P2P MIMO, wiretap, and multicasting channels. Then, existing solutions can be applied to obtain the precoding and power allocation matrices. The proposed PS can achieve near-capacity rate regions which are significantly higher compared to the existing orthogonal methods. In addition, in the case of the MIMO-NOMA networks without multicasting, a WSR maximization based on BSMM is formulated for all three scenarios. We generalize and prove that the zero duality gap holds for the WSR maximization, and the KKT conditions are necessary for the optimality. The two methods have their advantages. PS is a general tool for the MIMO-NOMA with different scenarios of security, while the WSR maximization provides a great potential for the secure MIMO-NOMA without multicasting. Both methods are computationally efficient compared with the DPC or S-DPC.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their suggestions to improve the quality of the paper.

REFERENCES

[1] Y. Qi and M. Vaezi, “Power splitting based precoding for the MIMO-BC with multicast and confidential messages,” in *Proc. IEEE Glob. Telecommun. Conf.*, 2020, pp. 1–6.
[2] Y. Qi and M. Vaezi, “Secure spectrum sharing in MIMO-NOMA,” in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw.*, 2021, pp. 213–220.
[3] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-orthogonal multiple access (NOMA) for cellular future radio access,” in *Proc. IEEE Veh. Technol. Conf.*, 2013, pp. 1–5.
[4] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Cham, Switzerland: Springer, 2019.
[5] Y. Cao, H. Sun, J. Soriaga, and T. Ji, “Resource spread multiple access-a novel transmission scheme for 5G uplink,” in *Proc. IEEE Veh. Technol. Conf.*, 2017, pp. 1–5.
[6] H. Weingarten, Y. Steinberg, and S. S. Shamai, “The capacity region of the gaussian multiple-input multiple-output broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
[7] Z. Ding, F. Adachi, and H. V. Poor, “The application of MIMO to non-orthogonal multiple access,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.
[8] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, “On the sum rate of MIMO-NOMA and MIMO-OMA systems,” *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 534–537, Aug. 2017.
[9] V.-D. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and O.-S. Shin, “Precoder design for signal superposition in MIMO-NOMA multicell networks,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2681–2695, Dec. 2017.
[10] Y. Huang, C. Zhang, J. Wang, Y. Jing, L. Yang, and X. You, “Signal processing for MIMO-NOMA: Present and future challenges,” *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 32–38, Apr. 2018.
[11] H. Weingarten, Y. Steinberg, and S. Shamai, “On the capacity region of the multi-antenna broadcast channel with common messages,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2006, pp. 2195–2199.
[12] M. Vaezi and H. V. Poor, “NOMA: An information-theoretic perspective,” in *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Cham, Switzerland: Springer, 2019, pp. 167–193.
[13] B. Clerckx *et al.*, “Is NOMA efficient in multi-antenna networks? A critical look at next generation multiple access techniques,” *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1310–1343, 2021.
[14] R. Liu, T. Liu, H. V. Poor, and S. Shamai, “Multiple-input multiple-output Gaussian broadcast channels with confidential messages,” *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4215–4227, Sep. 2010.
[15] R. Liu, T. Liu, H. V. Poor, and S. Shamai, “New results on multiple-input multiple-output broadcast channels with confidential messages,” *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1346–1359, Mar. 2013.
[16] R. F. Schaefer and H. Boche, “Physical layer service integration in wireless networks: Signal processing challenges,” *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 147–156, May 2014.
[17] A. D. Wyner, “The wire-tap channel,” *Bell System Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.
[18] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, “Transmit beamforming for physical-layer multicasting,” *IEEE Trans. Signal Process.*, vol. 54, no. 6-1, pp. 2239–2251, Jun. 2006.
[19] A. Khisti and G. W. Wornell, “Secure transmission with multiple antennas—Part II: The MIMOME wiretap channel,” *IEEE Trans. Inf. Theory*, vol. 11, no. 56, pp. 5515–5532, Nov. 2010.
[20] E. Ekrem and S. Ulukus, “On Gaussian MIMO broadcast channels with common and private messages,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2010, pp. 565–569.

[21] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2087–2104, Apr. 2014.

[22] H. D. Ly, T. Liu, and Y. Liang, "MIMO broadcasting with common, private and confidential messages," in *Proc. Int. Symp. Inf. Theory Appl*., 2008, pp. 1–6.

[23] Z. Goldfeld and H. H. Permuter, "MIMO Gaussian broadcast channels with common, private, and confidential messages," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2525–2544, Apr. 2019.

[24] E. Ekrem and S. Ulukus, "Capacity region of Gaussian MIMO broadcast channels with common and confidential messages," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5669–5680, Sep. 2012.

[25] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[26] J. Lee and N. Jindal, "High SNR analysis for MIMO broadcast channels: Dirty paper coding versus linear precoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4787–4792, Dec. 2007.

[27] H. Zhu, N. Prasad, and S. Rangarajan, "Precoder design for physical layer multicasting," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5932–5947, Nov. 2012.

[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. New York, NY, USA: Wiley, 2006.

[29] S. A. A. Fakoorian and A. L. Swindlehurst, "Optimal power allocation for GSVD-based beamforming in the MIMO Gaussian wiretap channel," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 2321–2325.

[30] 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 J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1714–1727, Sep. 2013.

[31] M. Vaezi, W. Shin, and H. V. Poor, "Optimal beamforming for Gaussian MIMO wiretap channels with two transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6726–6735, Oct. 2017.

[32] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[33] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. IEEE Int. Symp. Inf. Theory*, 2004, Art. no. 174.

[34] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 802–811, Jun. 2003.

[35] Z. Chen, Z. Ding, X. Dai, and R. Schober, "Asymptotic performance analysis of GSVD-NOMA systems with a large-scale antenna array," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 575–590, Jan. 2019.

[36] S. A. A. Fakoorian and A. L. Swindlehurst, "On the optimality of linear precoding for secrecy in the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1701–1713, Sep. 2013.

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

[38] Y. Qi and M. Vaezi, "Secure transmission in MIMO-NOMA networks," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2696–2700, Dec. 2020.

[39] H. Ly, T. Liu, and Y. Liang, "Multiple-input multiple-output Gaussian broadcast channels with common and confidential messages," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5477–5487, Sep. 2010.

[40] W. Mei, Z. Chen, and J. Fang, "GSVD-based precoding in MIMO systems with integrated services," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1528–1532, Nov. 2016.

[41] M. Vaezi, Y. Qi, and X. Zhang, "A rotation-based precoding for MIMO broadcast channels with integrated services," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1708–1712, Nov. 2019.

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

[43] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 2006.

[44] X. Zhang, Y. Qi, and M. Vaezi, "A rotation-based method for precoding in gaussian MIMOME channels," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1189–1200, Feb. 2021.

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

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

[47] W. Mei, Z. Chen, and J. Fang, "Secrecy capacity region maximization in Gaussian MISO channels with integrated services," *IEEE Signal Process. Lett.*, vol. 23, no. 8, pp. 1146–1150, Aug. 2016.

Yue Qi (Graduate Student Member, IEEE) received the B.Eng. and the M.Eng. degrees in electronic engineering from Xidian University, Xi'an, China, in 2015 and 2018, respectively. Since 2018, she has been working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, Villanova University, Villanova, PA, USA. Her research interests include NOMA, MIMO system, IoT heterogeneous networks, green communication, and signal processing. She was the recipient of the IEEE Communication Society Student Grant in ICC'20. She was a Session Chair of DySPAN'21.

Mojtaba Vaezi (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, and the Ph.D. degree in electrical engineering from McGill University, Montreal, QC, Canada,. From 2015 to 2018, he was with Princeton University, Princeton, NJ, USA, as a Postdoctoral Research Fellow and an Associate Research Scholar. He is currently an Assistant Professor of ECE with Villanova University, Villanova, PA, USA. Before joining Princeton, he was a Researcher with Ericsson Research in Montreal, Canada. He has authored or coauthored the book *Multiple Access Techniques for 5G Wireless Networks and Beyond* (Springer, 2019) in his research areas, which include signal processing and machine learning for wireless communications with an emphasis on physical layer security and fifth-generation (5G) and beyond radio access technologies.

Dr. Vaezi is the Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE COMMUNICATIONS LETTERS. He has co-organized six NOMA workshops at IEEE VTC 2017-Spring, Globecom'17, 18, and ICC'18, 19, 20. He was the recipient of several academic, leadership, and research awards, including the McGill Engineering Doctoral Award, IEEE Larry K. Wilson Regional Student Activities Award in 2013, Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship in 2014, Ministry of Science and ICT of Korea's Best Paper Award in 2017, IEEE Communications Letters Exemplary Editor Award in 2018, 2020 IEEE Communications Society Fred W. Ellersick Prize, and 2021 IEEE Philadelphia Section Delaware Valley Engineer of the Year Award.