Generalized Quadratic Matrix Programming: A Unified Framework for Linear Precoding With Arbitrary Input Distributions

Juening Jin, Yahong Rosa Zheng, *Fellow, IEEE*, Wen Chen, *Senior Member, IEEE*, and Chengshan Xiao, *Fellow, IEEE*

Abstract—This paper investigates a new class of nonconvex optimization, which provides a unified framework for linear precoding in single/multiuser multiple-input multiple-output channels with arbitrary input distributions. The new optimization is called generalized quadratic matrix programming (GQMP). Due to the nondeterministic polynomial time hardness of GQMP problems, instead of seeking globally optimal solutions, we propose an efficient algorithm that is guaranteed to converge to a Karush-Kuhn-Tucker point. The idea behind this algorithm is to construct explicit concave lower bounds for nonconvex objective and constraint functions, and then solve a sequence of concave maximization problems until convergence. In terms of application, we consider a downlink underlay secure cognitive radio network, where each node has multiple antennas. We design linear precoders to maximize the average secrecy (sum) rate with finite-alphabet inputs and statistical channel state information at the transmitter. The precoding problems under secure multicast/broadcast scenarios are GQMP problems, and thus, they can be solved efficiently by our proposed algorithm. Several numerical examples are provided to show the efficacy of our algorithm.

Index Terms—Generalized quadratic matrix programming, non-convex optimization, MIMO, linear precoding, secrecy sum rate maximization, arbitrary input distributions.

J. Jin was with Missouri University of Science and Technology, Rolla, MO 65409 USA. He is now with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: jueningjin@gmail.com).

Y. R. Zheng and C. Xiao are with the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409 USA (e-mail: zhengyr@mst.edu; xiaoc@mst.edu).

W. Chen is with the Shanghai Key Laboratory of Navigation and Location Based Services, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the School of Electronics Engineering and Automation, Guilin University of Electronics Technology, Guilin 541004, China (e-mail: wenchen@sjtu.edu.cn).

I. INTRODUCTION

OPTIMIZATION has been widely used in communications and signal processing. In many situations, the design and analysis of communication networks, when converted into mathematical forms, become certain types of optimization problems. However, finding the optimal solution to a general optimization problem is far from trivial. It is widely believed that the "watershed" in optimization is between convex and non-convex problems. Specifically, for any convex problem, the ellipsoid algorithm [1] can be used to get a global optimum with arbitrary precision, and its complexity is a polynomial function with the problem size. In contrast, non-convex problems are generally nondeterministic polynomial time (NP)-hard, which implies that there exists no polynomial time algorithm that can solve general non-convex problems to global optimality unless the complexity classes P and NP are proven to be equal.

Although it is challenging to handle non-convex optimization, much progress has been made for certain types of non-convex problems, such as difference-of-convex (DC) programming, quadratic constrained quadratic programming (QCQP) and signomial programming (SP), by means of convex optimization approaches [2]–[9]. In [2], the authors revisited a DC algorithm that can address DC problems, whose objective and constraint functions are the difference of two convex functions (not necessarily differentiable). When both objective and constraints of a DC problem are differentiable, the DC algorithm in [2] becomes another algorithm called convex-concave procedure [3], [4]. The work in [6] introduced the semidefinite relaxation (SDR) technique for non-convex QCQP problems. The main idea of SDR is to lift QCQP problems into the positive semidefinite matrix space, and then relax the non-convex rank one constraint. This technique is very powerful especially when the problem size is small. Moreover, a feasible point pursuit successive convex approximation was proposed in [7] for nonconvex QCQP problems. Finally, references [8], [9] proposed two different numerical algorithms, which are derived from the majorize-minimization (MM) framework [10], for non-convex SP problems.

However, many non-convex problems in communications and signal processing cannot be cast as DC, QCQP or SP problems. An important example is the linear precoder design for mutual information maximization in single-user multiple-input

multiple-output (MIMO) Gaussian channels with finite-alphabet inputs [11]–[15]. It was revealed in [15] that the input-output mutual information $\mathcal{I}(\mathbf{P}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{P})$ is the composition of a concave function $\mathcal{I}(\mathbf{W})$ and a quadratic matrix function $\mathbf{W} = \mathbf{P}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{P}$, where $\mathbf{P}$ is the precoding matrix. Such a composite function is neither convex nor concave with respect to $\mathbf{P}$, and it cannot be expressed as a DC, quadratic, or signomial function. For multiuser MIMO channels, including [16]–[21], precoding problems under finite-alphabet inputs are even more difficult, because they are generalizations of the single-user case. To the best of our knowledge, there does not exist any specific class of non-convex optimization that can capture the underlying structure of linear precoding for different communication channels and arbitrary channel input distributions.

A. Contributions

The contributions of this paper are listed as follows:

First, we study a new class of non-convex optimization, which provides a unified framework for linear precoding under different MIMO channels and arbitrary input distributions. The new optimization is a generalization of the quadratic matrix programming [22], and we call it generalized quadratic matrix programming (GQMP). A GQMP problem is defined as maximizing a generalized quadratic matrix function subject to generalized quadratic matrix inequality constraints, with both objective and constraints being non-convex functions. In this paper, we develop a numerical algorithm to solve GQMP problems efficiently. The solution obtained by our proposed GQMP algorithm reaches the Karsuh-Kuhn-Tucker (KKT) point. The key idea of this algorithm is to construct a concave lower bound for any generalized quadratic matrix function, and then replace the non-convex objective and constraint functions with the corresponding concave lower bounds. Subsequently, we solve a sequence of concave maximization problems until convergence. We further analyze the computational complexity of the GQMP algorithm and discuss two non-smooth generalizations of standard GQMP problems.

Second, we consider a downlink underlay secure cognitive radio (CR) network where a secondary-user transmitter (ST) communicates with I secondary-user receivers (SRs) in the presence of J eavesdroppers (EDs) and subject to interference threshold constraints at K primary-user receivers (PRs). Each node in the network is equipped with multiple antennas. We address the fundamental problem of maximizing the average secrecy (sum) rate of secondary users through linear precoding under the following assumptions: 1) The ST employs finite-alphabet modulation schemes; 2) The ST only has the knowledge of statistical channel state information (CSI) of each network node. The linear precoding problems under both secure multicast and secure broadcast scenarios are GQMP problems, thus they can be solved efficiently by our proposed GQMP algorithm. Finally, we present several numerical results to evaluate the performance of the proposed precoding with different system parameters. These results show that when considering finite-alphabet systems, our proposed precoding significantly outperforms the conventional Gaussian precoding design.

B. Notations

The following notations are adopted throughout the paper: Boldface lowercase letters, boldface uppercase letters, and calligraphic letters are used to denote vectors, matrices and sets, respectively. The real and complex number fields are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. The space of Hermitian $n \times n$ matrices is denoted by $\mathbb{H}^n$. The superscripts $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$ and $(\cdot)^{\mathrm{H}}$ stand for transpose, conjugate, and conjugate transpose operations, respectively. $\mathrm{tr}(\cdot)$ is the trace of a matrix; $[\cdot]^+$ denotes $\max(\cdot, 0)$; $\mathbf{dom}(\cdot)$ denotes the domain of a function; $\mathbf{A}^{(+)}$ denotes the positive definite part of a Hermitian matrix $\mathbf{A}$, i.e., $\mathbf{A}^{(+)} = \sum_{\lambda_i > 0} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{H}}$, where λ_i is the i-th eigenvalue of $\mathbf{A}$, and $\mathbf{u}_i$ is the corresponding eigenvector of $\mathbf{A}$; $\mathbf{A}^{(-)}$ denotes the negative definite part of $\mathbf{A}$, i.e., $\mathbf{A}^{(-)} = \sum_{\lambda_i < 0} \lambda_i \mathbf{u}_i \mathbf{u}_i^{\mathrm{H}}$; $\|\cdot\|$ denotes the Euclidean norm of a vector; $E_{\mathbf{x}}(\cdot)$ represents the statistical expectation with respect to $\mathbf{x}$; $\mathbf{I}$ and $\mathbf{0}$ denote an identity matrix and a zero matrix, respectively, with appropriate dimensions; $\mathbf{A} \succeq \mathbf{B}$ represents $\mathbf{A} - \mathbf{B}$ is positive semidefinite; $\mathcal{I}(\cdot)$ represents the mutual information; $\log(\cdot)$ and $\ln(\cdot)$ are used for the base two logarithm and natural logarithm, respectively.

The rest of this paper is organized as follows. Section II introduces and solves generalized quadratic matrix programming problems. Section III sets up the network model and formulates the precoding problems. Section IV solves the linear precoding problems by generalized quadratic matrix programming. Section IV presents several numerical results and Section V draws the conclusion.

II. GENERALIZED QUADRATIC MATRIX PROGRAMMING

A real-valued function $h(\mathbf{X})$ is said to be a composite quadratic matrix function if $h(\mathbf{X})$ can be expressed in the form

$$h(\mathbf{X}) = g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}), \quad g \in \mathcal{G} \tag{1}$$

where $\mathbf{X} \in \mathbb{C}^{n \times r}$, $\mathbf{A} \in \mathbb{H}^n$, $g(\mathbf{W}) : \mathbb{H}^r \to \mathbb{R}$, and $\mathcal{G}$ is the family of differentiable convex functions satisfying either matrix nondecreasing (MND) or matrix nonincreasing (MNI) condition. The definition of MND and MNI are given respectively as [23, ch. 3.6.1]:

$$\mathrm{MND}: \text{ if } \mathbf{W}_1 \succeq \mathbf{W}_2, \, g(\mathbf{W}_1) \geq g(\mathbf{W}_2) \tag{2}$$

$$\mathrm{MNI}: \text{ if } \mathbf{W}_1 \succeq \mathbf{W}_2, \, g(\mathbf{W}_1) \leq g(\mathbf{W}_2). \tag{3}$$

A linear combination of composite quadratic matrix functions is called a generalized quadratic matrix function

$$f(\mathbf{X}) = \sum_{k=1}^{K} \alpha_k g_k(\mathbf{X}^{\mathrm{H}}\mathbf{A}_k\mathbf{X}) \tag{4}$$

where $\alpha_k \in \mathbb{R}$, $k = 1, 2, ..., K$, $\mathbf{A}_k \in \mathbb{H}^n$, $k = 1, 2, ..., K$ and $g_k \in \mathcal{G}$, $k = 1, 2, ..., K$.

Programming problems dealing with generalized quadratic matrix functions are called generalized quadratic matrix programming (GQMP) problems. The standard form of a GQMP problem considered in this paper is given by

$$\begin{aligned} \underset{\mathbf{X} \in \mathcal{X}}{\mathrm{maximize}} \quad & f_0(\mathbf{X}) \\ \mathrm{subject\ to} \quad & f_j(\mathbf{X}) \geq 0, \, j = 1, 2, ..., J \end{aligned} \tag{5}$$

where $\mathbf{X} \in \mathbb{C}^{n \times r}$, $\mathcal{X}$ is a compact convex set, and $f_j(\mathbf{X})$, $j = 0, 1, ..., J$ are generalized quadratic matrix functions

$$f_j(\mathbf{X}) = \sum_{k=1}^{K_j} \alpha_{jk} g_{jk}(\mathbf{X}^H \mathbf{A}_{jk} \mathbf{X}). \tag{6}$$

Here we implicitly assume that the domain of problem (5) is an open set containing $\mathcal{X}$, i.e.,

$$\mathcal{X} \subset \mathbf{dom} f_0 \cap \mathbf{dom} f_1 \cap ... \cap \mathbf{dom} f_J. \tag{7}$$

This assumption ensures 1) $f_j(\mathbf{X})$, $j = 0, 1, ..., J$ are differentiable at every $\mathbf{X} \in \mathcal{X}$; 2) the domain $\bigcap_j \mathbf{dom} f_j$ has no effect on the optimal solution of problem (5).

Some major properties of the GQMP problem (5) are listed as follows:

1) The GQMP problem can be expressed in the minimization form, since this is equivalent to the maximization of $-f_0(\mathbf{X})$, which is again a generalized quadratic matrix function. Similarly, the constraints of (5) can be expressed as $f_j(\mathbf{X}) \leq 0$, $j = 1, 2, ..., J$.

2) The GQMP problem is a purely non-convex optimization problem, because generalized quadratic matrix functions $f_j(\mathbf{X})$, $j = 0, 1, ..., J$ are non-concave with respect to $\mathbf{X}$. When $J = 0$, the simplified problem $\underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \, f_0(\mathbf{X})$ is still a non-convex optimization problem.

3) When $\mathcal{X} = \mathbb{C}^{n \times r}$, and $g_{jk}(\mathbf{W}) = \text{tr}(\mathbf{W}) - c_{jk}$ for all (j, k), the GQMP problem is reduced to a non-convex quadratic matrix programming problem [22]

$$\begin{aligned} \underset{\mathbf{X} \in \mathbb{C}^{n \times r}}{\text{maximize}} \quad & \text{tr}(\mathbf{X}^H \tilde{\mathbf{A}}_0 \mathbf{X}) - \tilde{c}_0 \\ \text{subject to} \quad & \text{tr}(\mathbf{X}^H \tilde{\mathbf{A}}_j \mathbf{X}) - \tilde{c}_j \geq 0, \ j = 1, 2, ..., J \end{aligned} \tag{8}$$

where $\tilde{\mathbf{A}}_j = \sum_{k=1}^{K_j} \alpha_{jk} \mathbf{A}_{jk}$, $j = 0, 1, ..., J$, and $\tilde{c}_j = \sum_{k=1}^{K_j} \alpha_{jk} c_{jk}$, $j = 0, 1, ..., J$. Problem (8) can be solved by the SDR technique, i.e., define $\mathbf{Q} = \mathbf{X}\mathbf{X}^H$ and then relax the non-convex constraint $\text{rank}(\mathbf{Q}) \leq \min\{n, r\}$.

4) The GQMP problem belongs to the class of NP-hard problems, since the quadratic matrix programming problem in (8) is NP-hard in general.

Before stating our main results, we introduce a few definitions on the complex derivative and gradient. For a univariate function $f(x) : \mathbb{C} \to \mathbb{R}$, the definition of the complex derivative is given in [24]:

$$\frac{\partial f}{\partial x^*} \triangleq \frac{1}{2} \left(\frac{\partial f}{\partial \Re(x)} + j \frac{\partial f}{\partial \Im(x)} \right) \tag{9}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ are the real and image parts of a complex variable, respectively. For a multivariate function $f(\mathbf{X}) : \mathbb{C}^{n \times r} \to \mathbb{R}$, the complex gradient matrix $\nabla_{\mathbf{x}} f(\mathbf{X})$ is defined as

$$\nabla_{\mathbf{x}} f(\mathbf{X}) \triangleq \left[\frac{\partial f}{\partial \mathbf{X}_{ij}^*} \right] \tag{10}$$

where $\mathbf{X}_{ij}$ denotes the (i, j)-th element of $\mathbf{X}$.

A. Motivation

GQMP has a wide variety of applications in communications. In this subsection, we discuss some typical single/

multi-user MIMO Gaussian channels, for which linear precoding with arbitrary input distributions can be formulated as GQMP problems.

1) Single-User MIMO Gaussian Channels: The single-user MIMO Gaussian channel is modeled as [15]

$$\mathbf{y} = \mathbf{HPx} + \mathbf{n} \tag{11}$$

where $\mathbf{H} \in \mathbb{C}^{n \times r}$ is the complex channel matrix, $\mathbf{P} \in \mathbb{C}^{r \times r}$ is the linear precoder, $\mathbf{n} \in \mathbb{C}^{n \times 1}$ is the independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian noise with zero-mean and unit-variance, and $\mathbf{x} \in \mathbb{C}^{r \times 1}$ is the arbitrarily distributed channel input signal with zero-mean and covariance $E_{\mathbf{x}}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}$.

In Theorem 1 of [15], the authors presented three properties of the input-output mutual information $\mathcal{I}(\mathbf{x}; \mathbf{y})$ with arbitrarily distributed $\mathbf{x}$:

$\mathcal{I}(\mathbf{x}; \mathbf{y})$ is a function of $\mathbf{W} = \mathbf{P}^H \mathbf{H}^H \mathbf{HP}$

$$\nabla_{\mathbf{w}} \mathcal{I}(\mathbf{x}; \mathbf{y}) = \mathbf{\Phi} \tag{12}$$

$\mathcal{I}(\mathbf{x}; \mathbf{y})$ is a concave function with respect to $\mathbf{W}$

where $\mathbf{\Phi}$ is known as the minimum mean square matrix, i.e.,

$$\mathbf{\Phi} = E\left[(\mathbf{x} - E[\mathbf{x}|\mathbf{y}])(\mathbf{x} - E[\mathbf{x}|\mathbf{y}])^H \right]. \tag{13}$$

The first property shows that $\mathcal{I}(\mathbf{x}; \mathbf{y})$ is a function of $\mathbf{W}$, thus it can be expressed as $\mathcal{I}(\mathbf{W})$. The second property guarantees that $\mathcal{I}(\mathbf{W})$ is MND because $\mathbf{\Phi}$ is a positive semidefinite matrix [23, ch. 3.6.1]. The third property implies that $-\mathcal{I}(\mathbf{W})$ is a differentiable convex function of $\mathbf{W}$. Based on the definition in (4), $\mathcal{I}(\mathbf{P}^H \mathbf{H}^H \mathbf{HP})$ is a generalized quadratic matrix function of $\mathbf{P}$. Furthermore, since the feasible precoders with maximum transmit power γ form a convex set $\{\mathbf{P}|\text{tr}(\mathbf{P}^H \mathbf{P}) \leq \gamma\}$, the following mutual information maximization problem is a GQMP problem:

$$\begin{aligned} \underset{\mathbf{P}}{\text{maximize}} \quad & \mathcal{I}(\mathbf{P}^H \mathbf{H}^H \mathbf{HP}) \\ \text{subject to} \quad & \text{tr}(\mathbf{P}^H \mathbf{P}) \leq \gamma. \end{aligned} \tag{14}$$

Example 1: If $\mathbf{x}$ is complex Gaussian distributed, the input-output mutual information is given by

$$\mathcal{I}(\mathbf{x}; \mathbf{y}) = \log \det(\mathbf{I} + \mathbf{P}^H \mathbf{H}^H \mathbf{HP}). \tag{15}$$

Since $\mathcal{I}(\mathbf{W}) = \log \det(\mathbf{I} + \mathbf{W})$ is a concave and MND function with respect to $\mathbf{W}$, (15) is a generalized quadratic matrix function of $\mathbf{P}$.

Example 2: If each element of $\mathbf{x}$ is uniformly distributed from a Q-ary discrete constellation set, the input-output mutual information is given by [15]

$$\mathcal{I}(\mathbf{x}; \mathbf{y}) = r \log Q - \frac{1}{Q^r} \sum_{m=1}^{Q^r} E_{\mathbf{n}} \left\{ \log \sum_{k=1}^{Q^r} e^{-d_{m,k}} \right\}. \tag{16}$$

where $d_{m,k} = \|\mathbf{HP}(\mathbf{x}_m - \mathbf{x}_k) + \mathbf{n}\|^2 - \|\mathbf{n}\|^2$. According to (12), the mutual information expression in (16) is a generalized quadratic matrix function of $\mathbf{P}$.

2) MIMO Gaussian Broadcast Channels: The MIMO Gaussian broadcast channel is modeled as [17]

$$\mathbf{y}_i = \mathbf{H}_i \left(\sum_{j=1}^{m} \mathbf{P}_j \mathbf{x}_j \right) + \mathbf{n}_i, \quad i = 1, 2, ..., m \qquad (17)$$

where $\mathbf{H}_i \in \mathbb{C}^{n \times r}$ is the complex channel matrix for the i-th receiver, $\mathbf{n}_i \in \mathbb{C}^{n \times 1}$ is the i.i.d. circularly symmetric complex Gaussian noise with zero-mean and unit-variance; $\mathbf{P}_j \in \mathbb{C}^{r \times r}$ and $\mathbf{x}_j \in \mathbb{C}^{r \times 1}$ are the linear precoder and the channel input signal for the j-th receiver, respectively. We assume that $\{\mathbf{x}_j\}_{1 \leq j \leq m}$ are independent, and $\mathbf{x}_j \in \mathbb{C}^{r \times 1}$ is arbitrarily distributed with zero-mean and covariance $E_{\mathbf{x}_j}[\mathbf{x}_j \mathbf{x}_j^{\mathrm{H}}] = \mathbf{I}$.

The weighted sum-rate maximization and power minimization problems with linear precoding can be formulated respectively as

$$\begin{aligned} \underset{\mathbf{P}}{\text{maximize}} \quad & \sum_{i=1}^{m} \mu_i R_i(\mathbf{P}) \\ \text{subject to} \quad & \mathrm{tr}(\mathbf{P}^{\mathrm{H}}\mathbf{P}) \leq \gamma. \end{aligned} \qquad (18)$$

$$\begin{aligned} \underset{\mathbf{P}}{\text{minimize}} \quad & \mathrm{tr}(\mathbf{P}^{\mathrm{H}}\mathbf{P}) \\ \text{subject to} \quad & R_i(\mathbf{P}) \geq \bar{R}_i, \quad \forall i. \end{aligned} \qquad (19)$$

where $\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_m]$, $\mu_i \geq 0$ with $\sum_i \mu_i = m$, $R_i(\mathbf{P}) = \mathcal{I}(\mathbf{x}_i; \mathbf{y}_i)$ represents the achievable rate for the i-th receiver, γ is the maximum total transmit power, and $\bar{R}_i$ is the minimum rate requirement for the i-th receiver. Using the chain rule for mutual information [25], $R_i(\mathbf{P})$ can be expressed alternatively as

$$R_i(\mathbf{P}) = \mathcal{I}(\{\mathbf{x}_j\}_{1 \leq j \leq m}; \mathbf{y}_i) - \mathcal{I}(\{\mathbf{x}_j\}_{j \neq i}; \mathbf{y}_i | \mathbf{x}_i). \qquad (20)$$

According to (12), $R_i(\mathbf{P})$ is a generalized quadratic matrix function because it is the difference of two composite quadratic matrix functions. Thus problems (18) and (19) belong to the class of GQMP problems.

3) MIMO Gaussian Interference Channels: The MIMO Gaussian interference channel is modeled as [19]

$$\mathbf{y}_j = \mathbf{H}_{jj}\mathbf{P}_j\mathbf{x}_j + \sum_{i=1, i \neq j}^{m} \mathbf{H}_{ij}\mathbf{P}_i\mathbf{x}_i + \mathbf{n}_j, \quad j = 1, 2, ..., m \qquad (21)$$

where $\mathbf{H}_{ij} \in \mathbb{C}^{n \times r}$ is the complex channel matrix between the i-th transmitter and the j-th receiver, $\mathbf{n}_i \in \mathbb{C}^{n \times 1}$ is the i.i.d. circularly symmetric complex Gaussian noise with zero-mean and unit-variance; $\mathbf{P}_i \in \mathbb{C}^{r \times r}$ and $\mathbf{x}_i \in \mathbb{C}^{r \times 1}$ are the linear precoder and the channel input signal at the i-th transmitter, respectively. We assume that $\{\mathbf{x}_i\}_{1 \leq i \leq m}$ are independent, and $\mathbf{x}_i \in \mathbb{C}^{r \times 1}$ is arbitrarily distributed with zero-mean and covariance $E_{\mathbf{x}_i}[\mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}] = \mathbf{I}$.

The weighted sum-rate maximization problem with linear precoding can be formulated as

$$\begin{aligned} \underset{\{\mathbf{P}_i\}}{\text{maximize}} \quad & \sum_{j=1}^{m} \mu_j R_j(\{\mathbf{P}_i\}) \\ \text{subject to} \quad & \mathrm{tr}(\mathbf{P}_i^{\mathrm{H}}\mathbf{P}_i) \leq \gamma_i, \quad i = 1, 2, ..., m. \end{aligned} \qquad (22)$$

where $\mu_j \geq 0$ with $\sum_j \mu_j = m$, $\{\mathbf{P}_i\}$ denotes the collection of all precoders $\{\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_m\}$, $R_j(\{\mathbf{P}_i\}) = \mathcal{I}(\mathbf{x}_j; \mathbf{y}_j)$ represents the achievable rate at the j-th transmitter, and γ_i is the maximum transmit power for the i-th transmitter. Using the chain rule for mutual information, $R_j(\{\mathbf{P}_i\})$ can be expressed alternatively as

$$R_j(\{\mathbf{P}_i\}) = \mathcal{I}(\{\mathbf{x}_i\}_{1 \leq i \leq m}; \mathbf{y}_j) - \mathcal{I}(\{\mathbf{x}_i\}_{i \neq j}; \mathbf{y}_j | \mathbf{x}_j). \qquad (23)$$

The achievable rate $R_j(\{\mathbf{P}_i\})$ is the difference of two composite quadratic matrix functions with respect to $\mathbf{P} = \mathrm{diag}\{\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_m\}$, where $\mathrm{diag}\{\cdot\}$ represents a block diagonal matrix. Therefore, $R_j(\{\mathbf{P}_i\})$ is a generalized quadratic matrix function of $\mathbf{P}$, and problem (22) belongs to the class of GQMP problems.

4) MIMO Gaussian Wiretap Channels: The MIMO Gaussian wiretap channel is modeled as [16]

$$\begin{aligned} \mathbf{y}_{\mathrm{r}} &= \mathbf{H}_{\mathrm{r}}\mathbf{P}\mathbf{x} + \mathbf{n}_{\mathrm{r}} \\ \mathbf{y}_{\mathrm{e}} &= \mathbf{H}_{\mathrm{e}}\mathbf{P}\mathbf{x} + \mathbf{n}_{\mathrm{e}} \end{aligned} \qquad (24)$$

where $\mathbf{y}_{\mathrm{r}} \in \mathbb{C}^{n \times 1}$ and $\mathbf{y}_{\mathrm{e}} \in \mathbb{C}^{n \times 1}$ are received signals at the intended receiver and the eavesdropper, respectively; $\mathbf{H}_{\mathrm{r}} \in \mathbb{C}^{n \times r}$ and $\mathbf{H}_{\mathrm{e}} \in \mathbb{C}^{n \times r}$ are complex channel matrices; $\mathbf{P} \in \mathbb{C}^{r \times r}$ is the linear precoder at the transmitter; $\mathbf{n}_{\mathrm{r}} \in \mathbb{C}^{n \times 1}$ and $\mathbf{n}_{\mathrm{e}} \in \mathbb{C}^{n \times 1}$ are i.i.d. circularly symmetric complex Gaussian noises with zero-means and covariances $\sigma^2\mathbf{I}$; and $\mathbf{x} \in \mathbb{C}^{r \times 1}$ is the arbitrarily distributed channel input signal with zero-mean and covariance $E_{\mathbf{x}}[\mathbf{x}\mathbf{x}^{\mathrm{H}}] = \mathbf{I}$.

The secrecy rate maximization problem with linear precoding can be formulated as

$$\begin{aligned} \underset{\mathbf{P}}{\text{maximize}} \quad & \mathcal{I}(\mathbf{x}; \mathbf{y}_{\mathrm{r}}) - \mathcal{I}(\mathbf{x}; \mathbf{y}_{\mathrm{e}}) \\ \text{subject to} \quad & \mathrm{tr}(\mathbf{P}^{\mathrm{H}}\mathbf{P}) \leq \gamma. \end{aligned} \qquad (25)$$

where γ is the maximum transmit power. Since $\mathcal{I}(\mathbf{x}; \mathbf{y}_{\mathrm{r}}) - \mathcal{I}(\mathbf{x}; \mathbf{y}_{\mathrm{e}})$ is a generalized quadratic matrix function of $\mathbf{P}$, problem (25) belongs to the class of GQMP problems.

B. Algorithm Design

In this subsection, we design a numerical algorithm for GQMP problems by investigating the underlying structure of composite quadratic matrix functions $g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X})$. For every $g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X})$, we provide the corresponding concave lower bound $l(\mathbf{X}; \mathbf{X}_0)$ and convex upper bound $u(\mathbf{X}; \mathbf{X}_0)$, which depend on an arbitrary matrix $\mathbf{X}_0 \in \mathbb{C}^{n \times r}$ and satisfy the following three conditions:

$l(\mathbf{X}; \mathbf{X}_0) \leq g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}) \leq u(\mathbf{X}; \mathbf{X}_0)$ for all $\mathbf{X}$

$g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}) = l(\mathbf{X}; \mathbf{X}_0) = u(\mathbf{X}; \mathbf{X}_0)$ when $\mathbf{X} = \mathbf{X}_0$ $\qquad (26)$

$\nabla_{\mathbf{x}} g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}) = \nabla_{\mathbf{x}} l(\mathbf{X}; \mathbf{X}_0) = \nabla_{\mathbf{x}} u(\mathbf{X}; \mathbf{X}_0)$ when $\mathbf{X} = \mathbf{X}_0$.

In other words, $g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X})$ lies between $l(\mathbf{X}; \mathbf{X}_0)$ and $u(\mathbf{X}; \mathbf{X}_0)$, and it is tangent to both $l(\mathbf{X}; \mathbf{X}_0)$ and $u(\mathbf{X}; \mathbf{X}_0)$ when $\mathbf{X} = \mathbf{X}_0$. The conditions in (26) are necessary for us to design an ascent algorithm that converges to a KKT point of problem (5).

Algorithm 1: The GQMP algorithm.

1) Initialization: Given tolerance $\epsilon > 0$, choose an initial feasible point $\mathbf{X}_0$, set $n = 1$, $s_0 = f_0(\mathbf{X}_0)$. Let $\mathbf{X}_n^{\text{opt}}$ represents the optimal solution of (31) at the n-th iteration.
2) Stopping criterion: if $|s_n - s_{n-1}| > \epsilon$ go to the next step,
   otherwise STOP.
3) Concave approximation:
   a) replace $\mathbf{X}_0$ in (31) with $\mathbf{X}_n^{\text{opt}}$ and solve problem (31) to obtain $\mathbf{X}_{n+1}^{\text{opt}}$.
   b) set $s_{n+1} = f_0(\mathbf{X}_{n+1}^{\text{opt}})$.
4) Set $n := n + 1$ and go to step 2).
5) Output: $\mathbf{X}_n^{\text{opt}}$.

Theorem 1: The concave lower bound of $g(\mathbf{X}^{\mathrm{H}} \mathbf{A} \mathbf{X})$ is

$$l(\mathbf{X}; \mathbf{X}_0) = \mathrm{tr}\big(\mathbf{L}(\mathbf{X})^{\mathrm{H}} \mathbf{G}\big) + g(\mathbf{X}_0^{\mathrm{H}} \mathbf{A} \mathbf{X}_0) - \mathrm{tr}\big(\mathbf{X}_0^{\mathrm{H}} \mathbf{A} \mathbf{X}_0 \mathbf{G}\big) \tag{27}$$

where $\mathbf{G} \in \mathbb{H}^r$ is the complex gradient of $g(\mathbf{W})$ at $\mathbf{W} = \mathbf{X}_0^{\mathrm{H}} \mathbf{A} \mathbf{X}_0$, i.e., $\mathbf{G} = \nabla_{\mathbf{w}} \, g(\mathbf{X}_0^{\mathrm{H}} \mathbf{A} \mathbf{X}_0)$, and $\mathbf{L}(\mathbf{X})$ is given by

$$\mathbf{L}(\mathbf{X}) = \begin{cases} \mathbf{L}_1, & g(\mathbf{W}) \text{ is MND} \\ \mathbf{L}_2, & g(\mathbf{W}) \text{ is MNI} \end{cases} \tag{28}$$

with $\mathbf{L}_1 = \mathbf{X}^{\mathrm{H}} \mathbf{A}^{(-)} \mathbf{X} + \mathbf{X}^{\mathrm{H}} \mathbf{A}^{(+)} \mathbf{X}_0 + \mathbf{X}_0^{\mathrm{H}} \mathbf{A}^{(+)} \mathbf{X} - \mathbf{X}_0^{\mathrm{H}} \mathbf{A}^{(+)} \mathbf{X}_0$, $\mathbf{L}_2 = \mathbf{X}^{\mathrm{H}} \mathbf{A}^{(+)} \mathbf{X} + \mathbf{X}^{\mathrm{H}} \mathbf{A}^{(-)} \mathbf{X}_0 + \mathbf{X}_0^{\mathrm{H}} \mathbf{A}^{(-)} \mathbf{X} - \mathbf{X}_0^{\mathrm{H}} \mathbf{A}^{(-)} \mathbf{X}_0$.

Proof: See Appendix A. ∎

Theorem 2: The convex upper bound of $g(\mathbf{X}^{\mathrm{H}} \mathbf{A} \mathbf{X})$ is

$$u(\mathbf{X}; \mathbf{X}_0) = \begin{cases} g\big(\mathbf{U}_1(\mathbf{X})\big), & g(\mathbf{W}) \text{ is MND} \\ g\big(\mathbf{U}_2(\mathbf{X})\big), & g(\mathbf{W}) \text{ is MNI} \end{cases} \tag{29}$$

with

$$\mathbf{U}_1(\mathbf{X}) = \mathbf{X}^{\mathrm{H}} \mathbf{A}^{(+)} \mathbf{X} + \mathbf{X}^{\mathrm{H}} \mathbf{A}^{(-)} \mathbf{X}_0 + \mathbf{X}_0^{\mathrm{H}} \mathbf{A}^{(-)} \mathbf{X} - \mathbf{X}_0^{\mathrm{H}} \mathbf{A}^{(-)} \mathbf{X}_0$$

$$\mathbf{U}_2(\mathbf{X}) = \mathbf{X}^{\mathrm{H}} \mathbf{A}^{(-)} \mathbf{X} + \mathbf{X}^{\mathrm{H}} \mathbf{A}^{(+)} \mathbf{X}_0 + \mathbf{X}_0^{\mathrm{H}} \mathbf{A}^{(+)} \mathbf{X} - \mathbf{X}_0^{\mathrm{H}} \mathbf{A}^{(+)} \mathbf{X}_0.$$

Proof: See Appendix A. ∎

Based on Theorems 1 and 2, a concave lower bound of $f_j(\mathbf{X})$ in (6) is given by

$$\bar{f}_j(\mathbf{X}; \mathbf{X}_0) = \sum_{\alpha_{jk} > 0} \alpha_{jk} l_{jk}(\mathbf{X}; \mathbf{X}_0) + \sum_{\alpha_{jk} < 0} \alpha_{jk} u_{jk}(\mathbf{X}; \mathbf{X}_0) \tag{30}$$

where $l_{jk}(\mathbf{X}; \mathbf{X}_0)$ and $u_{jk}(\mathbf{X}; \mathbf{X}_0)$ represent the lower and upper bounds of $g_{jk}(\mathbf{X}^{\mathrm{H}} \mathbf{A}_{jk} \mathbf{X})$, respectively. By replacing each of $f_j(\mathbf{X})$, $j = 0, 1, ..., J$ with the corresponding concave lower bound $\bar{f}_j(\mathbf{X}; \mathbf{X}_0)$, we obtain the following concave maximization problem

$$\begin{array}{ll} \underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} & \bar{f}_0(\mathbf{X}; \mathbf{X}_0) \\ \text{subject to} & \bar{f}_j(\mathbf{X}; \mathbf{X}_0) \geq 0, \; j = 1, 2, ..., J. \end{array} \tag{31}$$

Then the MM framework [10] can be exploited to find a KKT point of problem (5) through solving a sequence of problems (31) with different $\mathbf{X}_0$. In the first iteration, we solve (31) at initial $\mathbf{X}_0$, and the optimal solution is denoted by $\mathbf{X}_1^{\text{opt}}$. Then

we replace $\bar{f}_j(\mathbf{X}; \mathbf{X}_0)$ in (31) with $\bar{f}_j(\mathbf{X}; \mathbf{X}_1^{\text{opt}})$, $j = 0, 1, ..., J$, and solve problem (31) again. At the n-th iteration, we solve (31) by replacing $\bar{f}_j(\mathbf{X}; \mathbf{X}_0)$ with $\bar{f}_j(\mathbf{X}; \mathbf{X}_{n-1}^{\text{opt}})$, $j = 0, 1, ..., J$, where $\mathbf{X}_{n-1}^{\text{opt}}$ is the optimal solution of (31) at the $(n-1)$-th iteration. The GQMP algorithm for solving problem (5) is summarized in Algorithm 1.

The convergence of Algorithm 1 is presented by the following proposition.

Proposition 1: Every limit point of the iterates $\{\mathbf{X}_n^{\text{opt}}\}$ generated by Algorithm 1 satisfies the KKT conditions of problem (5).

Proof: See Appendix B. ∎

Algorithm 1 is an iterative procedure between updating concave lower bounds and solving concave maximization problems (31). Since (31) is a convex optimization problem, it can be solved efficiently by the interior-point method with Newton iterations. The total number of optimization variables in problem (31) is nr, then the complexity order for solving (31) with Newton iterations is about $\mathcal{O}((nr)^3)$ [26]. Assuming that Algorithm 1 updates the concave lower bounds T times, the overall complexity is then given by $\mathcal{O}(T(nr)^3)$.

C. Non-Smooth Generalization

The GQMP algorithm developed in the last subsection can be used to handle non-smooth optimization problems. Herein, we discuss two possible non-smooth generalizations.

1) GQMP With Min-Rate Utility: For a multiuser system, we often adopt a utility function to measure the overall system performance. Thus it is of interest to consider the following GQMP problem with the min-rate utility:

$$\begin{array}{ll} R = \underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} & \underset{1 \leq j \leq L}{\min} \, f_j(\mathbf{X}) \\ \text{subject to} & f_j(\mathbf{X}) \geq 0, \; j = L + 1, 2, ..., J \end{array} \tag{32}$$

where R is the optimal value of problem (32), $\mathbf{X} \in \mathbb{C}^{n \times r}$, $\mathcal{X}$ is a compact convex set, and $f_j(\mathbf{X})$, $j = 1, 2, ..., J$ are generalized quadratic matrix functions

$$f_j(\mathbf{X}) = \sum_{k=1}^{K_j} \alpha_{jk} g_{jk}(\mathbf{X}^{\mathrm{H}} \mathbf{A}_{jk} \mathbf{X}). \tag{33}$$

Problem (32) is a non-smooth optimization problem, and our first step is to replace the non-smooth point-wise minimum operator with the log-sum-exp approximation via the following inequality [27], [28]

$$0 \leq \underset{1 \leq j \leq L}{\min} \, a_j - \frac{1}{\beta} \ln \sum_{j=1}^{L} e^{\beta a_j} \leq \frac{1}{|\beta|} \ln L, \quad \beta < 0 \tag{34}$$

which results in a smooth optimization problem

$$\begin{array}{ll} R_s(\beta) = \underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} & \frac{1}{\beta} \ln \sum_{j=1}^{L} e^{\beta f_j(\mathbf{X})} \\ \text{subject to} & f_j(\mathbf{X}) \geq 0, \; j = L + 1, 2, ..., J. \end{array} \tag{35}$$

Here $R_s(\beta)$ is the optimal value of (35). The relationship between problems (32) and (35) is revealed as follows

$$|R_s(\beta) - R| < \frac{1}{|\beta|} \ln L. \qquad (36)$$

Then we can obtain a KKT point of (35) through solving a sequence of maximization problems

$$\underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \quad \frac{1}{\beta} \ln \sum_{j=1}^{L} e^{\beta \bar{f}_j(\mathbf{X};\mathbf{X}_0)}$$
$$\text{subject to} \quad \bar{f}_j(\mathbf{X};\mathbf{X}_0) \geq 0, \ j = L+1, 2, ..., J \qquad (37)$$

where $\bar{f}_j(\mathbf{X};\mathbf{X}_0)$ is the concave lower bound of $f_j(\mathbf{X})$, $j = 1, 2, ..., J$. The objective function of problem (37) is concave because $-\ln \sum_j \exp(-f_j)$ is concave whenever f_j is concave for all j. Therefore, problem (37) is a smooth concave maximization problem, whose optimal value can be readily attained. Subsequently, we invoke Algorithm 1 to solve (37) multiple times with sufficiently large $|\beta|$ until convergence, and a suboptimal solution of problem (32) is obtained.

2) Secrecy Sum Rate Maximization: Consider the secrecy sum rate maximization problem in multiuser multi-eavesdropper networks, where each node is equipped with multiple antennas. The secrecy sum rate in such networks has a very complicated form:

$$\sum_{i=1}^{I} \min_{1 \leq j \leq J} \left[R_{ij} \right]^+ \qquad (38)$$

where $\left[R_{ij} \right]^+$ represents the individual secrecy rate between the i-th user and the j-th eavesdropper. Replacing each of R_{ij} with a generalized quadratic matrix function, we obtain the following generalized GQMP problems:

$$\underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \quad \sum_{i=1}^{I} \min_{1 \leq j \leq J} \left[f_{ij}(\mathbf{X}) \right]^+ \qquad (39)$$

where $\mathbf{X} \in \mathbb{C}^{n \times r}$, $\mathcal{X}$ is a compact convex set, and $f_{ij}(\mathbf{X}), \forall(i,j)$ are generalized quadratic matrix functions

$$f_{ij}(\mathbf{X}) = \sum_{k=1}^{K_{ij}} \alpha_{ijk} g_{ijk}(\mathbf{X}^{\mathrm{H}} \mathbf{A}_{ijk} \mathbf{X}). \qquad (40)$$

Problem (39) is extremely difficult to solve because of the non-concavity of $[\cdot]^+$. To see this, we first reformulate problem (39) as

$$\underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \quad \sum_{i=1}^{I} \left[\min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) \right]^+. \qquad (41)$$

Invoking Theorems 1 and 2, we can construct the concave lower bound of $f_{ij}(\mathbf{X})$, which is denoted by $\bar{f}_{ij}(\mathbf{X};\mathbf{X}_0)$. Moreover, the point-wise minimum of concave functions is concave [23], i.e., $\min_j \bar{f}_{ij}(\mathbf{X};\mathbf{X}_0)$ is a concave function. However, $\sum_i [\min_j \bar{f}_{ij}(\mathbf{X};\mathbf{X}_0)]^+$ is neither concave nor quasi-concave, due to the non-concavity of $[\cdot]^+$. Therefore, Algorithm 1 cannot be applied directly to solve problem (39). The following proposition shows that a "naive" method can be used to solve problem

(39) via solving $2^I - 1$ GQMP problems:

$$R_m = \underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \quad \sum_{i \in \mathcal{S}_m} \min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) \qquad (42)$$

where R_m is the optimal value of (42), and $\mathcal{S}_m$ represents the m-th non-empty subset of $\{1, 2, ..., I\}$.

Proposition 2: Let R denote the optimal value of problem (39), then we have

$$R = \max_{1 \leq m \leq 2^I - 1} R_m. \qquad (43)$$

Proof: See Appendix B. ∎

Although Proposition 2 provides a trackable way to solve problem (39), the complexity of this approach grows exponentially with respect to I. In the sequel, we design a more efficient algorithm for problem (39).

Our first step is to reformulate problem (41) as

$$\underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \quad \sum_{i=1}^{I} \max_{\lambda_i \in [0,1]} \left[\lambda_i \cdot \min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) \right]. \qquad (44)$$

Problems (41) and (44) are equivalent because

$$\left[\min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) \right]^+ = \max_{\lambda_i \in [0,1]} \left[\lambda_i \cdot \min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) \right]. \qquad (45)$$

The optimal λ_i of problem (44) depends on $\mathbf{X}$, and it has the following semi-closed form expression:

$$\lambda_i^*(\mathbf{X}) = \begin{cases} 1, & \min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) \geq 0 \\ 0, & \min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) < 0 \end{cases}, \quad i = 1, 2, ..., I. \qquad (46)$$

By plugging $\lambda_i^*(\mathbf{X})$ into (44), problem (39) is equivalent to the following optimization problem

$$\underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \quad F(\mathbf{X}) = \sum_{i=1}^{I} \left[\lambda_i^*(\mathbf{X}) \cdot \min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) \right]. \qquad (47)$$

Note that for any $\mathbf{X}_0 \in \mathcal{X}$, the following inequalities hold

$$F(\mathbf{X}) = \sum_{i=1}^{I} \max_{\lambda_i \in [0,1]} \left[\lambda_i \cdot \min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) \right] \qquad (48)$$

$$\geq \sum_{i=1}^{I} \lambda_i^*(\mathbf{X}_0) \cdot \min_{1 \leq j \leq J} f_{ij}(\mathbf{X}) \qquad (49)$$

$$\geq \sum_{i=1}^{I} \lambda_i^*(\mathbf{X}_0) \cdot \min_{1 \leq j \leq J} \bar{f}_{ij}(\mathbf{X};\mathbf{X}_0). \qquad (50)$$

where $\bar{f}_{ij}(\mathbf{X};\mathbf{X}_0)$ is the concave lower bound of $f_{ij}(\mathbf{X})$. Since $\lambda_i^*(\mathbf{X}_0) \geq 0$, equation (50) serves as the concave lower bound of $F(\mathbf{X})$. Therefore, by replacing $F(\mathbf{X})$ with its concave lower bound in (50), we obtain a concave maximization problem

$$\underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \quad \bar{F}(\mathbf{X};\mathbf{X}_0) = \sum_{i=1}^{I} \lambda_i^*(\mathbf{X}_0) \cdot \min_{1 \leq j \leq J} \bar{f}_{ij}(\mathbf{X};\mathbf{X}_0). \qquad (51)$$

Using the log-sum-exp approximation in (34), we can solve problem (51) efficiently through solving a smooth convex problem. Then we can obtain a suboptimal solution of problem (39)

Algorithm 2: The Generalized GQMP algorithm.

1) Initialization: Given tolerance $\epsilon > 0$, choose an initial feasible point $\mathbf{X}_0$, set $n = 1$, $s_0 = F(\mathbf{X}_0)$. Let $\mathbf{X}_n^{\text{opt}}$ represents the optimal solution of (51) at the n-th iteration.
2) Stopping criterion: if $|s_n - s_{n-1}| > \epsilon$ go to the next step,
   otherwise STOP.
3) Concave approximation:
   a) replace $\mathbf{X}_0$ in (51) with $\mathbf{X}_n^{\text{opt}}$ and solve problem (51) to obtain $\mathbf{X}_{n+1}^{\text{opt}}$.
   b) set $s_{n+1} = F(\mathbf{X}_{n+1}^{\text{opt}})$.
4) Set $n := n + 1$ and go to step 2).
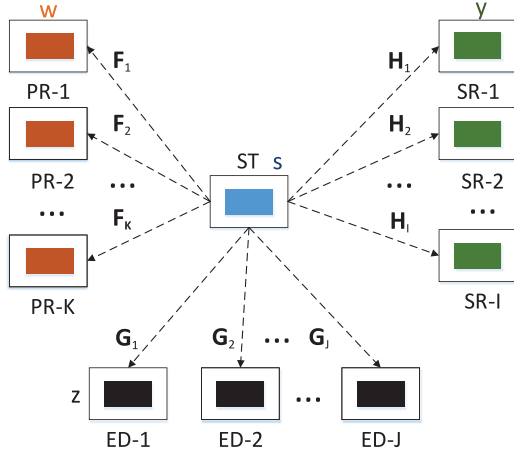5) Output: $\mathbf{X}_n^{\text{opt}}$.

Fig. 1. System model of a downlink underlay secure cognitive radio network.

by solving a sequence of problems (51) with different $\mathbf{X}_0$. The details are summarized in Algorithm 2. The convergence of Algorithm 2 for problem (39) is guaranteed by the following proposition.

Proposition 3: Let $\mathbf{X}_n^{\text{opt}}$ represent the optimal solution of problem (51) at the n-th iteration, then the sequence $\{F(\mathbf{X}_n^{\text{opt}})\}$ generated by Algorithm 2 converges.

Proof: See Appendix B. ∎

III. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a downlink underlay secure CR network depicted in Fig. 1. A secondary-user transmitter (ST) is serving I secondary-user receivers (SRs) in the presence of J eavesdroppers (EDs) and K primary-user receivers (PRs). The channel output at the i-th SR, the j-th ED and the k-th PR are, respectively, given by

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{s} + \mathbf{n}_{h_i}, \quad i = 1, 2, ..., I$$

$$\mathbf{z}_j = \mathbf{G}_j \mathbf{s} + \mathbf{n}_{g_j}, \quad j = 1, 2, ..., J$$

$$\mathbf{w}_k = \mathbf{F}_k \mathbf{s} + \mathbf{n}_{f_k}, \quad k = 1, 2, ..., K \quad (52)$$

where $\mathbf{H}_i \in \mathbb{C}^{N_R \times N_T}$, $\mathbf{G}_j \in \mathbb{C}^{N_E \times N_T}$ and $\mathbf{F}_k \in \mathbb{C}^{N_P \times N_T}$ are complex channel matrices from the ST to the i-th SR, the j-th ED, and the k-th PR, respectively; $\mathbf{s} \in \mathbb{C}^{N_T \times 1}$ is the channel input at the ST; $\mathbf{n}_{h_i} \in \mathbb{C}^{N_R \times 1}$, $\mathbf{n}_{g_j} \in \mathbb{C}^{N_E \times 1}$ and $\mathbf{n}_{f_k} \in \mathbb{C}^{N_P \times 1}$

are i.i.d. circularly symmetric complex Gaussian noises with zero-mean and unit-variance.

In this paper, we assume Kronecker correlation model, where the channel matrices can be written as [29]

$$\mathbf{H}_i = \mathbf{\Phi}_{h_i}^{\frac{1}{2}} \tilde{\mathbf{H}}_i \mathbf{\Theta}_{h_i}^{\frac{1}{2}}, \quad i = 1, 2, ..., I$$

$$\mathbf{G}_j = \mathbf{\Phi}_{g_j}^{\frac{1}{2}} \tilde{\mathbf{G}}_j \mathbf{\Theta}_{g_j}^{\frac{1}{2}}, \quad j = 1, 2, ..., J$$

$$\mathbf{F}_k = \mathbf{\Phi}_{f_k}^{\frac{1}{2}} \tilde{\mathbf{F}}_k \mathbf{\Theta}_{f_k}^{\frac{1}{2}}, \quad k = 1, 2, ..., K. \quad (53)$$

Here $\tilde{\mathbf{H}}_i \in \mathbb{C}^{N_R \times N_T}$, $\tilde{\mathbf{G}}_j \in \mathbb{C}^{N_E \times N_T}$ and $\tilde{\mathbf{F}}_k \in \mathbb{C}^{N_P \times N_T}$ are random matrices with i.i.d. zero-mean unit-variance complex Gaussian entries; $\mathbf{\Theta}_{h_i} \in \mathbb{C}^{N_T \times N_T}$, $\mathbf{\Theta}_{g_j} \in \mathbb{C}^{N_T \times N_T}$ and $\mathbf{\Theta}_{f_k} \in \mathbb{C}^{N_T \times N_T}$ are positive semidefinite transmit correlation matrices of $\mathbf{H}_i$, $\mathbf{G}_j$ and $\mathbf{F}_k$, respectively. $\mathbf{\Phi}_{h_i} \in \mathbb{C}^{N_R \times N_R}$, $\mathbf{\Phi}_{g_j} \in \mathbb{C}^{N_E \times N_E}$ and $\mathbf{\Phi}_{f_k} \in \mathbb{C}^{N_P \times N_P}$ are positive semidefinite receive correlation matrices of $\mathbf{H}_i$, $\mathbf{G}_j$ and $\mathbf{F}_k$, respectively. We further assume that the receive antennas at SRs, EDs and PRs are uncorrelated, i.e., $\mathbf{\Phi}_{h_i}$, $\mathbf{\Phi}_{g_j}$ and $\mathbf{\Phi}_{f_k}$ are all identity matrices. Then the correlation model in (53) becomes

$$\mathbf{H}_i = \tilde{\mathbf{H}}_i \mathbf{\Theta}_{h_i}^{\frac{1}{2}}, \quad i = 1, 2, ..., I$$

$$\mathbf{G}_j = \tilde{\mathbf{G}}_j \mathbf{\Theta}_{g_j}^{\frac{1}{2}}, \quad j = 1, 2, ..., J$$

$$\mathbf{F}_k = \tilde{\mathbf{F}}_k \mathbf{\Theta}_{f_k}^{\frac{1}{2}}, \quad k = 1, 2, ..., K. \quad (54)$$

Note that the above uncorrelated assumption are only used to avoid cumbersome expressions in precoding design. As we will see in Section IV, even when $\mathbf{\Phi}_{h_i}$, $\mathbf{\Phi}_{g_j}$ and $\mathbf{\Phi}_{f_k}$ are arbitrary positive semidefinite matrices, precoding designs can also be cast as GQMP problems.

In the sequel, we focus on two different transmission scenarios in the downlink underlay secure CR network: secure multicast scenario and secure broadcast scenario.

A. Secure Multicast Scenario

In the secure multicast scenario, the ST sends a common message to all SRs. Therefore, the channel input $\mathbf{s}$ can be represented as

$$\mathbf{s} = \mathbf{P}\mathbf{x} \quad (55)$$

where $\mathbf{P} \in \mathbb{C}^{N_T \times N_T}$ is the precoding matrix at the ST, and $\mathbf{x} \in \mathbb{C}^{N_T \times 1}$ is the input data vector for SRs with zero-mean and covariance $E_{\mathbf{x}}[\mathbf{x}\mathbf{x}^{\text{H}}] = \mathbf{I}$. We further assume that the i-th SR knows the instantaneous channel realization of $\mathbf{H}_i$, the j-th ED knows the instantaneous channel realization of $\mathbf{G}_j$, and the ST only has channel statistics of all nodes in the system, i.e., the transmit correlation matrices $\{\mathbf{\Theta}_{h_i}, \mathbf{\Theta}_{g_j}, \mathbf{\Theta}_{f_k}, \forall(i, j, k)\}$ as well as the distributions of $\tilde{\mathbf{H}}_i$, $\tilde{\mathbf{G}}_j$ and $\tilde{\mathbf{F}}_k$. Then the average secrecy rate for the i-th SR can be expressed as

$$R_i = \min_{1 \leq j \leq J} \left[\mathcal{I}(\mathbf{x}_i; \mathbf{y}_i | \mathbf{H}_i) - \mathcal{I}(\mathbf{x}_i; \mathbf{z}_j | \mathbf{G}_j) \right]^+ \quad (56)$$

$$= \min_{1 \leq j \leq J} \left[E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}_i; \mathbf{y}_i | \mathbf{H}_i = \bar{\mathbf{H}}_i) \right.$$
$$\left. - E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}_i; \mathbf{z}_j | \mathbf{G}_j = \bar{\mathbf{G}}_j) \right]^+ \quad (57)$$

where $\bar{\mathbf{H}}_i$ and $\bar{\mathbf{G}}_j$ are given channel realizations of $\mathbf{H}_i$ and $\mathbf{G}_j$, respectively. For notational simplicity, in the following, we omit the given channel realization condition in mutual information expressions. Based on (57), the following average secrecy rate for multicast scenario is achievable [30]

$$R_{\mathrm{MC}}(\mathbf{P}) = \min_{\substack{1 \le i \le I \\ 1 \le j \le J}} \left[E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}; \mathbf{y}_i) - E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}; \mathbf{z}_j) \right]^+$$

$$= \left[\min_{\substack{1 \le i \le I \\ 1 \le j \le J}} \left[E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}; \mathbf{y}_i) - E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}; \mathbf{z}_j) \right] \right]^+. \quad (58)$$

We maximize $R_{\mathrm{MC}}(\mathbf{P})$ under the power constraint at the ST and the interference threshold constraints at PRs. The average transmit power at the ST is constrained to γ_0:

$$E_{\mathbf{x}} \mathrm{tr}\left(\mathbf{P}\mathbf{x}\mathbf{x}^{\mathrm{H}}\mathbf{P}^{\mathrm{H}}\right) = \mathrm{tr}(\mathbf{P}^{\mathrm{H}}\mathbf{P}) \le \gamma_0 \quad (59)$$

and the average interference power at the k-th PR is limited by $N_T \cdot \gamma_k$:

$$E_{\mathbf{x}, \mathbf{F}_k} \mathrm{tr}\left(\mathbf{F}_k \mathbf{P}\mathbf{x}\mathbf{x}^{\mathrm{H}}\mathbf{P}^{\mathrm{H}}\mathbf{F}_k^{\mathrm{H}}\right)$$

$$= E_{\tilde{\mathbf{F}}_k} \mathrm{tr}\left(\mathbf{P}^{\mathrm{H}}(\boldsymbol{\Theta}_{f_k}^{\frac{1}{2}})^{\mathrm{H}}\tilde{\mathbf{F}}_k^{\mathrm{H}}\tilde{\mathbf{F}}_k \boldsymbol{\Theta}_{f_k}^{\frac{1}{2}}\mathbf{P}\right)$$

$$= N_T \cdot \mathrm{tr}\left(\mathbf{P}^{\mathrm{H}}\boldsymbol{\Theta}_{f_k}\mathbf{P}\right) \le N_T \cdot \gamma_k, \ \forall k. \quad (60)$$

The second equality in (60) holds because 1) the entries of $\tilde{\mathbf{F}}_k$ are i.i.d. complex Gaussian variables with zero-mean and unit-variance; 2) $\tilde{\mathbf{F}}_k$ and $\mathbf{x}$ are independent. Since both (59) and (60) are quadratic matrix constraints, we can incorporate (59) into (60) by defining $\boldsymbol{\Theta}_{f_0} = \mathbf{I}$. Then the set of all feasible precoding matrices is given by

$$\mathcal{P}_{\mathrm{MC}} = \{\mathbf{P} | \mathrm{tr}\left(\mathbf{P}^{\mathrm{H}}\boldsymbol{\Theta}_{f_k}\mathbf{P}\right) \le \gamma_k, \ k = 0, 1, ..., K\} \quad (61)$$

and the linear precoding problem under secure multicast scenario can be formulated as

$$\underset{\mathbf{P} \in \mathcal{P}_{\mathrm{MC}}}{\text{maximize}} \quad \min_{\substack{1 \le i \le I \\ 1 \le j \le J}} \left[E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}; \mathbf{y}_i) - E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}; \mathbf{z}_j) \right]. \quad (62)$$

Here we remove the nonnegative operator $[\cdot]^+$ in $R_{\mathrm{MC}}(\mathbf{P})$ without loss of optimality because 1) the maximum average secrecy rate obtained by problem (62) is nonnegative since we can always achieve a zero objective value with $\mathbf{P} = \mathbf{0}$; 2) when the maximum average secrecy rate is positive, $[\cdot]^+$ has no effect on the optimal precoders.

B. Secure Broadcast Scenario

In the secure broadcast scenario, the ST sends a private message to each SR. Therefore, the channel input $\mathbf{s}$ can be represented as

$$\mathbf{s} = \sum_{i=1}^{I} \mathbf{P}_i \mathbf{x}_i = \mathbf{P}\mathbf{x} \quad (63)$$

where $\mathbf{P}_i \in \mathbb{C}^{N_T \times N_T}$ is the precoding matrix for the i-th SR, $\mathbf{x}_i \in \mathbb{C}^{N_T \times 1}$ is the input data vector for the i-th SR with zero-mean and covariance $E_{\mathbf{x}_i}[\mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}] = \mathbf{I}$, $\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_I]$ and $\mathbf{x} = [\mathbf{x}_1^{\mathrm{H}}, \mathbf{x}_2^{\mathrm{H}}, ..., \mathbf{x}_I^{\mathrm{H}}]^{\mathrm{H}}$. We further assume that the i-th SR knows the instantaneous channel realization of $\mathbf{H}_i$, the j-th

ED knows the instantaneous channel realization of $\mathbf{G}_j$, and the ST only has channel statistics of all nodes in the system. In addition, the i-th SR treats signals of other SRs as interference, and the j-th ED can at best decode the signal of the i-th SR while treating signals of other SRs as interference. Under these assumptions, the average secrecy rate for the i-th SR can be expressed as

$$R_i = \min_{1 \le j \le J} \left[E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}_i; \mathbf{y}_i) - E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}_i; \mathbf{z}_j) \right]^+. \quad (64)$$

Based on (64), the following average secrecy sum rate for broadcast scenario is achievable

$$R_{\mathrm{BC}}(\mathbf{P}) = \sum_{i=1}^{I} \min_{1 \le j \le J} \left[E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}_i; \mathbf{y}_i) - E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}_i; \mathbf{z}_j) \right]^+. \quad (65)$$

The average secrecy sum rate maximization in secure broadcast scenario with power and interference threshold constraints can then be formulated as

$$\underset{\mathbf{P} \in \mathcal{P}_{\mathrm{BC}}}{\text{maximize}} \quad R_{\mathrm{BC}}(\mathbf{P}) \quad (66)$$

where the feasible set $\mathcal{P}_{\mathrm{BC}}$ can be obtained from (59) and (60) with $\boldsymbol{\Theta}_{f_0} = \mathbf{I}$:

$$\mathcal{P}_{\mathrm{BC}} = \{\mathbf{P} | \mathrm{tr}\left(\mathbf{P}^{\mathrm{H}}\boldsymbol{\Theta}_{f_k}\mathbf{P}\right) \le \gamma_k, \ k = 0, 1, ..., K\}. \quad (67)$$

IV. PRECODING DESIGN BY GENERALIZED QUADRATIC MATRIX PROGRAMMING

A. Secure Multicast Scenario

In this subsection, we solve problem (62) under finite-alphabet inputs by GQMP. Instead of Gaussian inputs, we assume that the input data vector $\mathbf{x}$ is uniformly distributed from $\mathcal{X}^{N_T} = \{\mathbf{x} | x_k \in \mathcal{X}, \forall k\}$, where x_k is the k-th element of $\mathbf{x}$, and $\mathcal{X}$ is a Q-ary discrete constellation set. Then the average constellation-constrained mutual information $E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}; \mathbf{y}_i)$ and $E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}; \mathbf{z}_j)$ can be expressed as [15]

$$E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}; \mathbf{y}_i) = \log M - \frac{1}{M} \sum_{m=1}^{M} E_{\mathbf{H}_i, \mathbf{n}_{h_i}} \left\{ \log \sum_{n=1}^{M} e^{-h_{m,n}^i} \right\} \quad (68)$$

$$E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}; \mathbf{z}_j) = \log M - \frac{1}{M} \sum_{m=1}^{M} E_{\mathbf{G}_j, \mathbf{n}_{g_j}} \left\{ \log \sum_{n=1}^{M} e^{-g_{m,n}^j} \right\} \quad (69)$$

where M is equal to Q^{N_T}; $h_{m,n}^i = \|\mathbf{H}_i \mathbf{P}(\mathbf{x}_m - \mathbf{x}_n) + \mathbf{n}_{h_i}\|^2 - \|\mathbf{n}_{h_i}\|^2$ and $g_{m,n}^j = \|\mathbf{G}_j \mathbf{P}(\mathbf{x}_m - \mathbf{x}_n) + \mathbf{n}_{g_j}\|^2 - \|\mathbf{n}_{g_j}\|^2$, with $\mathbf{x}_m$ and $\mathbf{x}_n$ representing input realizations from $\mathcal{X}^{N_T}$.

The average constellation-constrained mutual information is difficult to compute directly because both $E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}; \mathbf{y}_i)$ and $E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}; \mathbf{z}_j)$ in (68) and (69) have no closed form expressions. Moreover, the gradients of $E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}; \mathbf{y}_i)$ and $E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}; \mathbf{z}_j)$ also have no closed form expressions. Although we can use Monte Carlo method and numerical integral to estimate $E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}; \mathbf{y}_i)$ and $E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}; \mathbf{z}_j)$ as well as their gradients, the computational complexity is prohibitively high especially when the dimensions of $\mathbf{H}_i$ and $\mathbf{G}_j$ are large.

This difficulty can be mitigated by introducing an accurate approximation of the average mutual information in doubly correlated fading channels [31]. The average secrecy rate with finite-alphabet inputs can then be approximated as

$$\left[\min_{\substack{1 \le i \le I \\ 1 \le j \le J}} \left[g(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{g_j} \mathbf{P}; N_E) - g(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{h_i} \mathbf{P}; N_R) \right] \right]^+ \quad (70)$$

where $g(\mathbf{W}; N)$ with $N \ge 1$ is given below

$$g(\mathbf{W}; N) = \frac{1}{M} \sum_{m=1}^{M} \log \sum_{n=1}^{M} \left(1 + \frac{1}{2} \mathbf{e}_{mn}^{\mathrm{H}} \mathbf{W} \mathbf{e}_{mn} \right)^{-N}. \quad (71)$$

The approximation in (70) is very accurate for arbitrary correlation matrices and precoders, and the computational complexity of (70) is several orders of magnitude lower than that of the original average secrecy rate [31].

Using (70) as an alterative to replace the average secrecy rate, problem (62) can be approximated as

$$\underset{\mathbf{P} \in \mathcal{P}_{\mathrm{MC}}}{\text{maximize}} \quad \min_{\substack{1 \le i \le I \\ 1 \le j \le J}} g(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{g_j} \mathbf{P}; N_E) - g(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{h_i} \mathbf{P}; N_R). \quad (72)$$

The following proposition indicates that (72) is a GQMP problem with the min-rate utility.

Proposition 4: $g(\mathbf{W}; N)$ is a convex and MNI function of $\mathbf{W}$.

Proof: See Appendix B. ∎

Based on Proposition 4, $g(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{g_j} \mathbf{P}; N_E) - g(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{h_i} \mathbf{P}; N_R)$ is a generalized quadratic matrix function of $\mathbf{P}$ for all (i, j). Therefore, (72) is a special case of problem (32) and we can solve it efficiently by Algorithm 1.

B. Secure Broadcast Scenario

In this subsection, we solve problem (66) under finite-alphabet inputs by GQMP. Instead of Gaussian inputs, we assume that the input data vector $\mathbf{x}$ is uniformly distributed from $\mathcal{X}^{N_T I} = \{\mathbf{x} | x_k \in \mathcal{X}, \forall k\}$, where x_k is the k-th element of $\mathbf{x}$, and $\mathcal{X}$ is a Q-ary equiprobable discrete constellation set. Then the average constellation-constrained mutual information $E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}_i; \mathbf{y}_i)$ and $E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}_i; \mathbf{z}_j)$ can be expressed as

$$E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}_i; \mathbf{y}_i) = \frac{1}{M} \sum_{m=1}^{M} E_{\mathbf{H}_i, \mathbf{n}_{h_i}} \left\{ \log(a_m^{ii}) \right\} \quad (73)$$

$$E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}_i; \mathbf{z}_j) = \frac{1}{M} \sum_{m=1}^{M} E_{\mathbf{G}_j, \mathbf{n}_{g_j}} \left\{ \log(b_m^{ij}) \right\} \quad (74)$$

with

$$a_m^{ii} = \frac{\sum_{n=1}^{M} e^{-\|\mathbf{H}_i \mathbf{P} \mathbf{I}_i (\mathbf{x}_m - \mathbf{x}_n) + \mathbf{n}_{h_i}\|^2 + \|\mathbf{n}_{h_i}\|^2}}{\sum_{n=1}^{M} e^{-\|\mathbf{H}_i \mathbf{P} (\mathbf{x}_m - \mathbf{x}_n) + \mathbf{n}_{h_i}\|^2 + \|\mathbf{n}_{h_i}\|^2}} \quad (75)$$

$$b_m^{ij} = \frac{\sum_{n=1}^{M} e^{-\|\mathbf{G}_j \mathbf{P} \mathbf{I}_i (\mathbf{x}_m - \mathbf{x}_n) + \mathbf{n}_{g_j}\|^2 + \|\mathbf{n}_{g_j}\|^2}}{\sum_{n=1}^{M} e^{-\|\mathbf{G}_j \mathbf{P} (\mathbf{x}_m - \mathbf{x}_n) + \mathbf{n}_{g_j}\|^2 + \|\mathbf{n}_{g_j}\|^2}} \quad (76)$$

where M is equal to $Q^{N_T I}$; $\mathbf{I}_i$ is a block diagonal matrix formed by replacing the i-th $N_T \times N_T$ block diagonal entry of the $N_T I \times N_T I$ identity matrix $\mathbf{I}$ with $\mathbf{0}$; $\mathbf{x}_m$ and $\mathbf{x}_n$ are input realizations from $\mathcal{X}^{N_T I}$.

Similarly, by adopting the accurate approximation in [31], the average secrecy sum rate with finite-alphabet inputs can be approximated respectively as

$$\sum_{i=1}^{I} \min_{1 \le j \le J} \left[\mathcal{I}_A(\mathbf{x}_i; \mathbf{y}_i) - \mathcal{I}_A(\mathbf{x}_i; \mathbf{z}_j) \right]^+ \quad (77)$$

where $\mathcal{I}_A(\mathbf{x}_i; \mathbf{y}_i)$ and $\mathcal{I}_A(\mathbf{x}_i; \mathbf{z}_j)$ are accurate approximations of $E_{\mathbf{H}_i} \mathcal{I}(\mathbf{x}_i; \mathbf{y}_i)$ and $E_{\mathbf{G}_j} \mathcal{I}(\mathbf{x}_i; \mathbf{z}_j)$ respectively:

$$\mathcal{I}_A(\mathbf{x}_i; \mathbf{y}_i) = g_i(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{h_i} \mathbf{P}; N_R) - g(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{h_i} \mathbf{P}; N_R) \quad (78)$$

$$\mathcal{I}_A(\mathbf{x}_i; \mathbf{z}_j) = g_i(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{g_j} \mathbf{P}; N_E) - g(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{g_j} \mathbf{P}; N_E) \quad (79)$$

with

$$g_i(\mathbf{W}; N) = \frac{1}{M} \sum_{m=1}^{M} \log \sum_{n=1}^{M} \left(1 + \frac{1}{2} \mathbf{e}_{mn}^{\mathrm{H}} \mathbf{I}_i^{\mathrm{H}} \mathbf{W} \mathbf{I}_i \mathbf{e}_{mn} \right)^{-N} \quad (80)$$

$$g(\mathbf{W}; N) = \frac{1}{M} \sum_{m=1}^{M} \log \sum_{n=1}^{M} \left(1 + \frac{1}{2} \mathbf{e}_{mn}^{\mathrm{H}} \mathbf{W} \mathbf{e}_{mn} \right)^{-N}. \quad (81)$$

Using (77) as an alternative, problem (66) can be approximated as

$$\underset{\mathbf{P} \in \mathcal{P}_{\mathrm{BC}}}{\text{maximize}} \quad \sum_{i=1}^{I} \min_{1 \le j \le J} \left[\mathcal{I}_A(\mathbf{x}_i; \mathbf{y}_i) - \mathcal{I}_A(\mathbf{x}_i; \mathbf{z}_j) \right]^+ \quad (82)$$

According to Proposition 4, $g_i(\mathbf{W}; N)$ and $g(\mathbf{W}; N)$ are convex and MNI functions of $\mathbf{W}$. Therefore, problem (82) is a special case of problem (39), and we can solve it efficiently by Algorithm 2.

C. Discussions

In this subsection, we show that precoding problems under the doubly correlated model (53) are GQMP problems. For convenience, we restrict our attention to the secure multicast scenario, and it is straightforward to extend our results to the secure broadcast scenario.

Our first step is to replace the average secrecy rate $R_{\mathrm{MC}}(\mathbf{P})$ by the following accurate approximation:

$$\min_{\substack{1 \le i \le I \\ 1 \le j \le J}} \left[\bar{g}(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{g_j} \mathbf{P}; \lambda(\boldsymbol{\Phi}_{g_j})) - \bar{g}(\mathbf{P}^{\mathrm{H}} \boldsymbol{\Theta}_{h_i} \mathbf{P}; \lambda(\boldsymbol{\Phi}_{h_i})) \right]^+ \quad (83)$$

where $\lambda(\cdot)$ represents the eigenvalues of a positive semidefinite matrix; $\bar{g}(\mathbf{W}; \mathbf{r})$ is given below [31]

$$\bar{g}(\mathbf{W}; \mathbf{r}) = \frac{1}{M} \sum_{m=1}^{M} \log \sum_{n=1}^{M} \prod_{q} \left(1 + \frac{r_q}{2} \mathbf{e}_{mn}^{\mathrm{H}} \mathbf{W} \mathbf{e}_{mn} \right)^{-1} \quad (84)$$

with $r_q \ge 0$ being the q-th element of $\mathbf{r}$. When the receive correlation matrix is an identity matrix $\mathbf{I}$, $\lambda(\mathbf{I})$ is the vector with all entries one, and $\bar{g}(\mathbf{W}; \lambda(\mathbf{I}))$ is reduced to $g(\mathbf{W}; N)$ in (71). Although $\bar{g}(\mathbf{W}; \mathbf{r})$ looks more complicated than $g(\mathbf{W}; N)$, the convexity and MNI property of $\bar{g}(\mathbf{W}; \mathbf{r})$ can be proved in the same manner as in Proposition 4. This leads to the following proposition.

Proposition 5: $\bar{g}(\mathbf{W}; \mathbf{r})$ is a convex and MNI function of $\mathbf{W}$.

Proof: The proof is omitted for brevity. ∎

The second step is to determine the interference threshold constraints at PRs. When $\mathbf{\Phi}_{f_k} \neq \mathbf{I}$, the average interference power at the k-th PR is given by

$$E_{\mathbf{x}, \mathbf{F}_k} \operatorname{tr}\left(\mathbf{F}_k \mathbf{P} \mathbf{x} \mathbf{x}^{\mathrm{H}} \mathbf{P}^{\mathrm{H}} \mathbf{F}_k^{\mathrm{H}}\right)$$
$$= E_{\tilde{\mathbf{F}}_k} \operatorname{tr}\left(\mathbf{P}^{\mathrm{H}}(\mathbf{\Theta}_{f_k}^{\frac{1}{2}})^{\mathrm{H}} \tilde{\mathbf{F}}_k^{\mathrm{H}} \mathbf{\Phi}_{f_k} \tilde{\mathbf{F}}_k \mathbf{\Theta}_{f_k}^{\frac{1}{2}} \mathbf{P}\right)$$
$$= \operatorname{tr}(\mathbf{\Phi}_{f_k}) \cdot \operatorname{tr}\left(\mathbf{P}^{\mathrm{H}} \mathbf{\Theta}_{f_k} \mathbf{P}\right). \quad (85)$$

Since $\operatorname{tr}(\mathbf{\Phi}_{f_k}) > 0$ and $\mathbf{\Theta}_{f_k} \succeq \mathbf{0}$, the average interference power is a convex function of $\mathbf{P}$.

According to Proposition 5, the approximated average secrecy rate is a generalized quadratic matrix function with the min-rate utility. In addition, the feasible set with power and interference threshold constraints is convex. Therefore, the precoding problem under secure multicast scenario is a GQMP problem, and it can be solved efficiently by Algorithm 1.

V. NUMERICAL RESULTS

In this section, we provide numerical examples to show that the GQMP algorithm for precoding design is numerically robust and computationally effective under various situations. For illustration purpose, we adopt the exponential correlation model

$$\left[\mathbf{R}(\rho)\right]_{ij} = \rho^{|i-j|}, \quad \rho \in [0, 1) \quad (86)$$

where ρ is the correlation coefficient between different antennas.

A. Example 1: Point-to-Point MIMO Gaussian Channel

In this example, we compare our proposed precoding with the globally optimal precoding algorithm [15] in a point-to-point MIMO setting. Specifically, we solve problem (14) with finite-alphabet inputs by the GQMP and the globally optimal precoding algorithms. The channel matrix $\mathbf{H}$ is given by

$$\mathbf{H} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \quad (87)$$

which was also used in [15]. The input signal is drawn from QPSK constellation. Since we assume unit noise power, the signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = \gamma$, where γ is the total transmit power at the transmitter.

Fig. 2 depicts comparison results with the globally optimal precoding algorithm [15] and the waterfilling algorithm. As shown in Fig. 2, our proposed precoding can achieve the performance of the globally optimal precoding algorithm in whole SNR regimes. In addition, the waterfilling algorithm, which is optimal for Gaussian inputs, is quite suboptimal for practical MIMO systems under finite-alphabet inputs, especially in the medium and high SNR regimes.

B. Example 2: MIMO Gaussian Wiretap Channel

In this example, we compare our proposed precoding with the gradient descent algorithm [16] in a MIMO wiretap channel. Specifically, we solve problem (25) with finite-alphabet inputs by the GQMP and the gradient descent algorithms. The channel matrices at the receiver and the eavesdropper are given
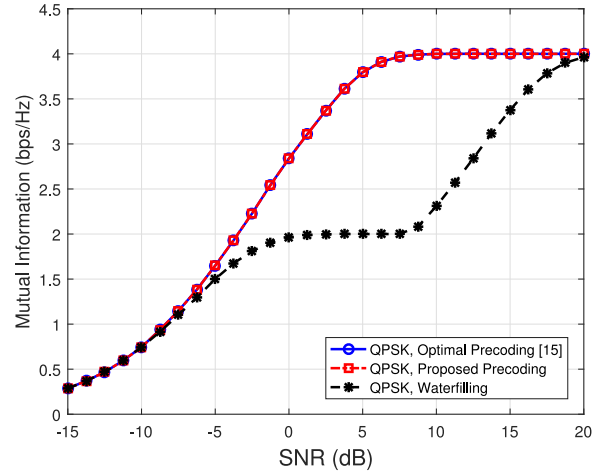

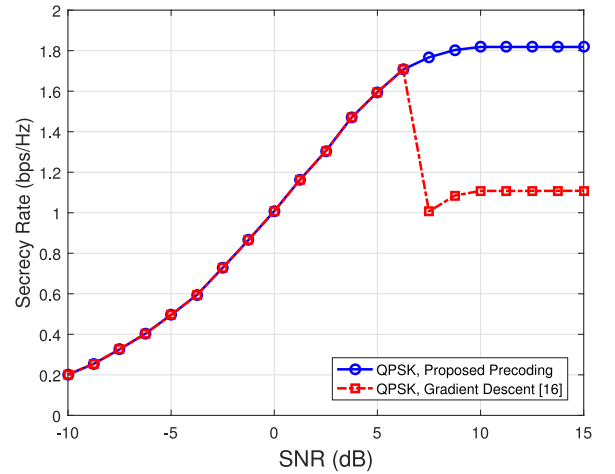Fig. 2. Mutual information as a function of the SNR.

Fig. 3. Comparison between the proposed precoding and the gradient descent algorithms with finite-alphabet inputs.

respectively as

$$\mathbf{H}_{\mathrm{r}} = \begin{bmatrix} 0.0991 - 0.8676i & 1.0814 + 1.1281i \end{bmatrix}, \quad (88)$$

$$\mathbf{H}_{\mathrm{e}} = \begin{bmatrix} 0.3880 + 1.2024i & -0.9825 + 0.5914i \\ 0.4709 - 0.3073i & 0.6815 - 0.2125i \end{bmatrix} \quad (89)$$

which was also used in [16]. The input signal is drawn from QPSK constellation, and the SNR is defined as $\mathrm{SNR} = \frac{\gamma}{\sigma^2}$, where σ^2 is the noise power at the receiver and the eavesdropper. In addition, the initial point for both algorithms is set as

$$\mathbf{P}_0 = \begin{bmatrix} 0.0312 - 0.1762i & 0.1719 + 0.7560i \\ 0.9126 + 0.5724i & -0.1064 - 0.0097i \end{bmatrix}. \quad (90)$$

Fig. 3 compares the secrecy rate performance of our proposed precoding with the gradient descent algorithm [16]. We observe that our algorithm outperforms the gradient descent algorithm in high SNR regimes because the gradient descent algorithm is susceptible to initial points.

To further show the influence of initial points for the proposed precoding and the gradient descent algorithms, we run these two algorithms 2000 times with Gaussian distributed initial points (applying a normalization to satisfy the power constraint).

TABLE I
DISTRIBUTION OF SECRECY RATE FOR VARIOUS INITIAL POINTS

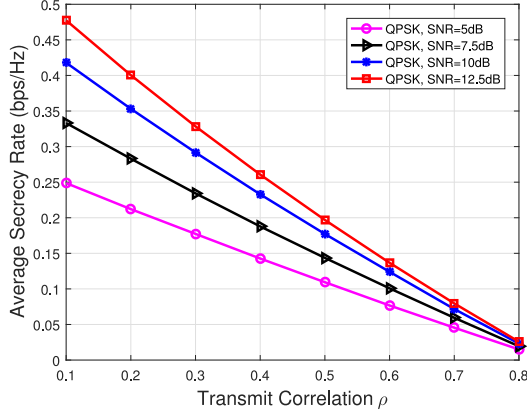| Secrecy Rate	Proposed Precoding	Gradient Descent
1.7471 bps/Hz	1831/2000	932/2000
1.0063 bps/Hz	169/2000	1068/2000

Fig. 4.   Average secrecy rate as a function of the ED's transmit correlation ρ.

Table I shows the distribution of secrecy rate for the proposed precoding and the gradient descent algorithm with $\mathrm{SNR} = 7.5$ dB. The result demonstrates the superiority of the proposed algorithm. Our algorithm has over 90% probability to converge to a better solution while the gradient descent algorithm only has 46.6% probability to achieve this goal.

C. Example 3: Secure Multicast Scenario With One ED

In this example, we consider a fading secure CR network with one ST, two SRs, one ED and one PR. Each node in the system has two antennas. The transmit correlation matrices are given by

$$\boldsymbol{\Theta}_{h_1} = \mathbf{R}(0.95), \boldsymbol{\Theta}_{h_2} = \mathbf{R}(0.85),$$
$$\boldsymbol{\Theta}_{g_1} = \mathbf{R}(\rho), \boldsymbol{\Theta}_{f_1} = \mathbf{R}(0.50). \tag{91}$$

The SNR is defined as $\mathrm{SNR} = \gamma_0$, where γ_0 is the total transmit power at the ST. The normalized interference threshold at the PR is 10 dB less than the total transmit power, i.e., $\gamma_1 = 0.1\gamma_0$.

Fig. 4 investigates the maximum average secrecy rate as a function of the ED's transmit correlation ρ with $\mathrm{SNR} = 5$ dB, 7.5 dB, 10 dB and 12.5 dB. The input signal is drawn from QPSK constellation. As we can see, the average secrecy rate is monotonically decreasing with respect to ρ. This interesting phenomenon occurs because the impact of transmit correlation depends on the channel knowledge [32]. Reference [32] considered a single-user channel and showed that if the transmitter only knows statistical CSI, the average capacity is Schur-convex with respect to the channel correlation, i.e., the more correlated the transmit antennas are, the more capacity can be achieved. Therefore, when ρ increases, we need more redundancy rate to confuse the ED and the average secrecy rate decreases. In the extreme case that $\rho = 0.85$, the average secrecy rate is always zero because $\boldsymbol{\Theta}_{h_2} = \boldsymbol{\Theta}_{g_1}$.

In Fig. 5, we investigate the average secrecy rate as a function of the SNR under BPSK, QPSK, 8PSK and 16QAM modula-
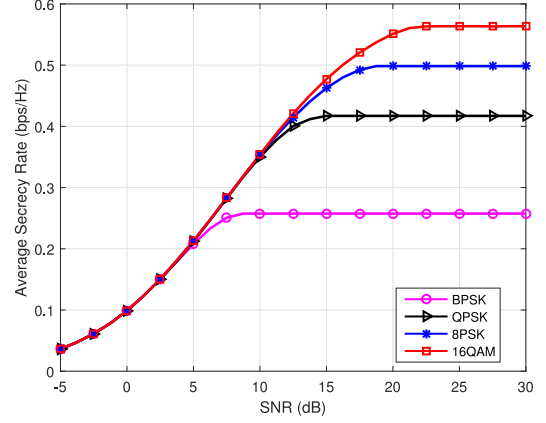

Fig. 5.   Average secrecy rate as a function of the SNR.

tions. The ED's transmit correlation is set as $\rho = 0.2$. Fig. 5 shows that precoding design by the proposed GQMP algorithm is very effective because it can achieve robust performances for a large-SNR range with various modulations. These results also indicate that we should adaptively determine the modulation based on the SNR. If the system works in the low SNR regime, we should use low-order modulations to reduce the complexity for precoder design. If the system works in the high SNR regime, we can use high-order modulations to achieve a better performance.

D. Example 4: Secure Multicast Scenario With Multiple EDs

In this example, we consider a fading secure CR network with one ST, two SRs, two EDs and two PRs. Each node in the system has two antennas. The transmit correlation matrices are given by

$$\boldsymbol{\Theta}_{h_1} = \mathbf{R}(0.9), \boldsymbol{\Theta}_{h_2} = \mathbf{R}(0.95), \boldsymbol{\Theta}_{g_1} = \mathbf{R}(0.3)$$
$$\boldsymbol{\Theta}_{g_2} = \mathbf{R}(0.4), \boldsymbol{\Theta}_{f_1} = \mathbf{R}(0.5), \boldsymbol{\Theta}_{f_2} = \mathbf{R}(0.7). \tag{92}$$

The normalized interference thresholds at PRs are 20 dB less than the total transmit power, i.e., $\gamma_1 = \gamma_2 = 0.01\gamma_0$, and the SNR is defined as $\mathrm{SNR} = \gamma_0$.

Fig. 6 depicts the comparison result with the Gaussian precoding design under QPSK and 16QAM modulations. The Gaussian precoding design employs Gaussian inputs to solve problem (62), where $E_{\mathbf{H}_i}\mathcal{I}(\mathbf{x}; \mathbf{y}_i)$ and $E_{\mathbf{G}_j}\mathcal{I}(\mathbf{x}; \mathbf{z}_j)$ with Gaussian inputs can be expressed as

$$E_{\mathbf{H}_i}\mathcal{I}(\mathbf{x}; \mathbf{y}_i) = E_{\mathbf{H}_i}\left[\log\det\left(\mathbf{I} + \mathbf{H}_i\mathbf{P}\mathbf{P}^{\mathrm{H}}\mathbf{H}_i^{\mathrm{H}}\right)\right] \tag{93}$$

$$E_{\mathbf{G}_j}\mathcal{I}(\mathbf{x}; \mathbf{z}_j) = E_{\mathbf{G}_j}\left[\log\det\left(\mathbf{I} + \mathbf{G}_j\mathbf{P}\mathbf{P}^{\mathrm{H}}\mathbf{G}_j^{\mathrm{H}}\right)\right]. \tag{94}$$

Let $\mathbf{Q} = \mathbf{P}\mathbf{P}^{\mathrm{H}}$, then the Gaussian precoding problem is given by

$$
\begin{aligned}
\underset{\mathbf{Q} \succeq \mathbf{0}}{\text{maximize}} \quad & \min_{\substack{1 \le i \le I \\ 1 \le j \le J}} \left[E_{\mathbf{H}_i}\mathcal{I}(\mathbf{x}; \mathbf{y}_i) - E_{\mathbf{G}_j}\mathcal{I}(\mathbf{x}; \mathbf{z}_j)\right] \\
\text{subject to} \quad & \operatorname{tr}(\boldsymbol{\Theta}_{f_k}\mathbf{Q}) \le \gamma_k, \ k = 0, 1, ..., K
\end{aligned}
\tag{95}
$$

Since problem (95) is a DC problem, we can obtain a locally optimal solution $\mathbf{Q}_{\mathrm{opt}}$ by the convex-concave procedure [3].

Fig. 6.   Proposed precoding versus Gaussian precoding design under different modulations, for secure multicast scenarios.
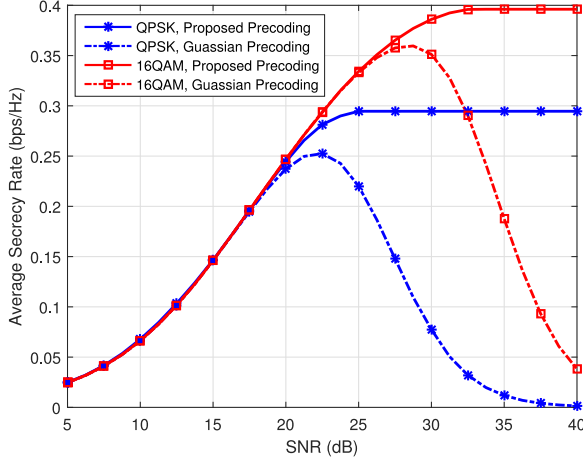

Fig. 7.   Proposed precoding versus Gaussian precoding design under different modulations, for secure broadcast scenarios.

After that, we evaluate the finite-alphabet based average secrecy rate under $\mathbf{P} = \mathbf{Q}_{\mathrm{opt}}^{\frac{1}{2}}$.

Based on the results in Fig. 6, we have the following remarks:

1) In the low SNR regime, finite-alphabet precoding and Gaussian precoding have the same performance. According to [14], the low-SNR expansion of mutual information is irrelevant to the input distribution, thus the optimal precoders designed under Gaussian inputs are also optimal for finite-alphabet inputs.

2) In the high SNR regime, the performance of Gaussian precoding design degrades severely with the increasing SNR. The reason is that both $E_{\mathbf{H}_i}\mathcal{I}(\mathbf{x}; \mathbf{y}_i)$ and $E_{\mathbf{G}_j}\mathcal{I}(\mathbf{x}; \mathbf{z}_j)$ in (68), (69) saturate at $\log M$ in the high SNR regime. Therefore, if we do not carefully control the transmit power $\mathrm{tr}(\mathbf{P}^{\mathrm{H}}\mathbf{P})$ in the high SNR, the average secrecy rate under finite-alphabet inputs will be $\log M - \log M = 0$. Since the objective function of (95) does not have this saturation property, the covariance matrix $\mathbf{Q}_{\mathrm{opt}}$ designed by problem (95) uses too much transmit power. Therefore, the corresponding average secrecy rate with finite-alphabet inputs degrades severely in the high SNR regime.

E. Example 5: Secure Broadcast Scenario With Multiple EDs

In this example, we consider a fading secure CR network with one ST, two SRs, two EDs and two PRs. Each node in the system has two antennas. The transmit correlation matrices are given by

$$\mathbf{\Theta}_{h_1} = \mathbf{R}(0.9), \mathbf{\Theta}_{h_2} = \mathbf{R}(0.8), \mathbf{\Theta}_{g_1} = \mathbf{R}(0.45),$$
$$\mathbf{\Theta}_{g_2} = \mathbf{R}(0.55), \mathbf{\Theta}_{f_1} = \mathbf{R}(0.2), \mathbf{\Theta}_{f_2} = \mathbf{R}(0.6). \quad (96)$$

The normalized interference thresholds at PRs are 10 dB less than the total transmit power, i.e., $\gamma_1 = \gamma_2 = 0.1\gamma_0$, and the SNR is defined as $\mathrm{SNR} = \gamma_0$.

Fig. 7 depicts the performance comparison with the Gaussian precoding design under BPSK and QPSK modulations. The Gaussian precoding solves problem (66) under Gaussian inputs by GQMP, and then evaluates the finite-alphabet based average secrecy sum rate with the corresponding suboptimal precoders. Results in Fig. 6 indicate that our proposed finite-alphabet pre-
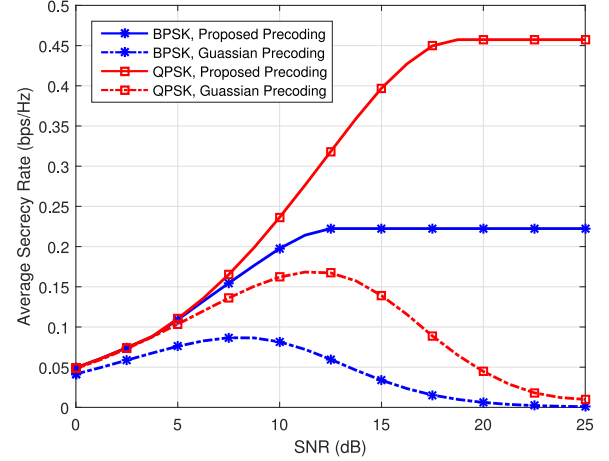
coding offers much higher secrecy sum rate than the Gaussian precoding in the medium and high SNR regimes. This is because the precoders $[\mathbf{P}_1, \mathbf{P}_2]$ designed by Gaussian precoding have the following form:

$$[\mathbf{P}_1, \mathbf{P}_2] = [\mathbf{P}_1^{\mathrm{opt}}, \mathbf{0}]. \quad (97)$$

Equation (97) implies that the Gaussian precoding design allocates all the power to the first SR, and the precoder for the second SR is $\mathbf{0}$. In contrast, our proposed finite-alphabet precoding allocates power to both SRs, and the precoded signal for the second SR acts as the jamming signal to further confuse EDs. Therefore, our proposed precoding significantly outperforms the Gaussian precoding design.

VI. Conclusion

In this paper, we have proposed the generalized quadratic matrix programming, which is a significant generalization of quadratic matrix programming. GQMP captures the inherent structure of input-output mutual information with arbitrary input distributions, thus it unifies the design of linear precoding under various MIMO Gaussian channels. By exploiting the features of generalized quadratic matrix functions, we have designed a low complexity algorithm that converges to the KKT point of any GQMP problem. Next, we have applied GQMP to design linear precoders in downlink underlay secure CR networks. The considered linear precoding target finite-alphabet inputs directly and exploit statistical CSI of fading channels. We have demonstrated that the precoding problems under both secure multicast and secure broadcast scenarios are GQMP problems, thus we can solve them efficiently by our proposed GQMP algorithm. Numerical results have shown that the proposed algorithm is both robust and effective.

APPENDIX A
PROOFS OF THEOREMS 1 AND 2

Proof of Theorem 1: Since $g(\mathbf{W})$ is a differentiable convex function, it can be lower bounded by its first-order Taylor ap-

proximation at any point $\mathbf{W}_0$ [23, ch. 3.1.3]:

$$g(\mathbf{W}) \geq g(\mathbf{W}_0) + \operatorname{tr}\big[(\mathbf{W} - \mathbf{W}_0)^{\mathrm{H}} \nabla_{\mathrm{w}}\, g(\mathbf{W}_0)\big]. \tag{98}$$

By plugging $\mathbf{W} = \mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}$, $\mathbf{W}_0 = \mathbf{X}_0^{\mathrm{H}}\mathbf{A}\mathbf{X}_0$ and $\mathbf{G} = \nabla_{\mathrm{w}}\, g(\mathbf{W}_0)$ into (98), we have

$$g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}) \geq \operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}\mathbf{G}) + g(\mathbf{X}_0^{\mathrm{H}}\mathbf{A}\mathbf{X}_0) - \operatorname{tr}(\mathbf{X}_0^{\mathrm{H}}\mathbf{A}\mathbf{X}_0\mathbf{G}). \tag{99}$$

Since $\operatorname{tr}(\mathbf{Y}^{\mathrm{H}}\mathbf{V}\mathbf{Y}\mathbf{Z}) = \operatorname{vec}(\mathbf{Y})^{\mathrm{H}} \cdot (\mathbf{Z}^{\mathrm{T}} \otimes \mathbf{V}) \cdot \operatorname{vec}(\mathbf{Y})$ [33], we can express $\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}\mathbf{G})$ alternatively as

$$\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X}\mathbf{G}) = \operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X}\mathbf{G}) + \operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X}\mathbf{G}) \tag{100}$$

$$= \operatorname{vec}(\mathbf{X})^{\mathrm{H}} \cdot \left[(\mathbf{G}^{\mathrm{T}} \otimes \mathbf{A}^{(+)}) + (\mathbf{G}^{\mathrm{T}} \otimes \mathbf{A}^{(-)})\right] \cdot \operatorname{vec}(\mathbf{X}). \tag{101}$$

If $g(\mathbf{W})$ is MND, its complex gradient $\mathbf{G} \succeq \mathbf{0}$ for all $\mathbf{X}_0$ [23, ch. 3.6.1]. Therefore, $\mathbf{G}^{\mathrm{T}} \otimes \mathbf{A}^{(+)} \succeq \mathbf{0}$ and $\mathbf{G}^{\mathrm{T}} \otimes \mathbf{A}^{(-)} \preceq \mathbf{0}$ [33], which imply the convexity of $\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X}\mathbf{G})$ and the concavity of $\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X}\mathbf{G})$ respectively. The convex function $\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X}\mathbf{G})$ can be further lower bounded by its first-order Taylor approximation at $\mathbf{X}_0$:

$$\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X}\mathbf{G})$$

$$\geq \operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X}_0 + \mathbf{X}_0^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X} - \mathbf{X}_0^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X}_0\mathbf{G}). \tag{102}$$

Combining (99), (100) and (102), we obtain a concave lower bound when $g(\mathbf{W})$ is MND.

If $g(\mathbf{W})$ is MNI, its complex gradient $\mathbf{G} \preceq \mathbf{0}$ for all $\mathbf{X}_0$ [23, ch. 3.6.1]. Therefore, $\mathbf{G}^{\mathrm{T}} \otimes \mathbf{A}^{(+)} \preceq \mathbf{0}$ and $\mathbf{G}^{\mathrm{T}} \otimes \mathbf{A}^{(-)} \succeq \mathbf{0}$ [33], which imply the concavity of $\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X}\mathbf{G})$ and the convexity of $\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X}\mathbf{G})$ respectively. The convex function $\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X}\mathbf{G})$ can be further lower bounded by its first-order Taylor approximation at $\mathbf{X}_0$:

$$\operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X}\mathbf{G})$$

$$\geq \operatorname{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X}_0 + \mathbf{X}_0^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X} - \mathbf{X}_0^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X}_0\mathbf{G}). \tag{103}$$

Combining (99), (100) and (103), we obtain a concave lower bound when $g(\mathbf{W})$ is MNI. ■

Proof of Theorem 2: Since $(\mathbf{X} - \mathbf{X}_0)^{\mathrm{H}}\mathbf{A}^{(-)}(\mathbf{X} - \mathbf{X}_0) \preceq \mathbf{0}$ and $(\mathbf{X} - \mathbf{X}_0)^{\mathrm{H}}\mathbf{A}^{(+)}(\mathbf{X} - \mathbf{X}_0) \succeq \mathbf{0}$, the following inequalities hold

$$\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X} \preceq \mathbf{U}_1(\mathbf{X}) \tag{104}$$

$$\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X} \succeq \mathbf{U}_2(\mathbf{X}) \tag{105}$$

with

$$\mathbf{U}_1(\mathbf{X}) = \mathbf{X}^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X} + \mathbf{X}^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X}_0 + \mathbf{X}_0^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X} - \mathbf{X}_0^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X}_0$$

$$\mathbf{U}_2(\mathbf{X}) = \mathbf{X}^{\mathrm{H}}\mathbf{A}^{(-)}\mathbf{X} + \mathbf{X}^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X}_0 + \mathbf{X}_0^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X} - \mathbf{X}_0^{\mathrm{H}}\mathbf{A}^{(+)}\mathbf{X}_0$$

Here $\mathbf{U}_1(\mathbf{X})$ is a convex function in the sense that for all $\mathbf{X}_1$, $\mathbf{X}_2$ and θ with $0 \leq \theta \leq 1$, we have

$$\mathbf{U}_1(\theta\mathbf{X}_1 + (1-\theta)\mathbf{X}_2) \preceq \theta\mathbf{U}_1(\mathbf{X}_1) + (1-\theta)\mathbf{U}_1(\mathbf{X}_2). \tag{106}$$

Based on the definition in (106), $-\mathbf{U}_2(\mathbf{X})$ is convex, thus $\mathbf{U}_2(\mathbf{X})$ is a concave function.

If $g(\mathbf{W})$ is MND, $g(\mathbf{U}_1(\mathbf{X}))$ serves as an upper bound of $g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X})$. Furthermore, for all $\mathbf{X}_1$, $\mathbf{X}_2$ and θ with $0 \leq \theta \leq 1$, we have

$$g(\mathbf{U}_1(\theta\mathbf{X}_1 + (1-\theta)\mathbf{X}_2))$$

$$\leq g(\theta\mathbf{U}_1(\mathbf{X}_1) + (1-\theta)\mathbf{U}_1(\mathbf{X}_2)) \tag{107}$$

$$\leq \theta g(\mathbf{U}_1(\mathbf{X}_1)) + (1-\theta)g(\mathbf{U}_1(\mathbf{X}_2)) \tag{108}$$

where (107) and (108) hold due to the MND condition and the convexity of $g(\mathbf{W})$, respectively. Therefore, $g(\mathbf{U}_1(\mathbf{X}))$ serves as a convex upper bound of $g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X})$ when $g(\mathbf{W})$ is MND.

If $g(\mathbf{W})$ is MNI, $g(\mathbf{U}_2(\mathbf{X}))$ serves as an upper bound of $g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X})$. Furthermore, for all $\mathbf{X}_1$, $\mathbf{X}_2$ and θ with $0 \leq \theta \leq 1$, we have

$$g(\mathbf{U}_2(\theta\mathbf{X}_1 + (1-\theta)\mathbf{X}_2))$$

$$\leq g(\theta\mathbf{U}_2(\mathbf{X}_1) + (1-\theta)\mathbf{U}_2(\mathbf{X}_2)) \tag{109}$$

$$\leq \theta g(\mathbf{U}_2(\mathbf{X}_1)) + (1-\theta)g(\mathbf{U}_2(\mathbf{X}_2)) \tag{110}$$

where (109) and (110) hold due to the MNI condition and the convexity of $g(\mathbf{W})$, respectively. Therefore, $g(\mathbf{U}_2(\mathbf{X}))$ serves as a convex upper bound of $g(\mathbf{X}^{\mathrm{H}}\mathbf{A}\mathbf{X})$ when $g(\mathbf{W})$ is MNI. ■

APPENDIX B
PROOFS OF PROPOSITIONS 1–4

Proof of Proposition 1: First, by introducing an auxiliary variable t, problem (5) is equivalent to

$$\begin{aligned} \underset{\mathbf{X}\in\mathcal{X},t}{\text{maximize}} \quad & t \\ \text{subject to} \quad & f_0(\mathbf{X}) - t \geq 0 \\ & f_j(\mathbf{X}) \geq 0, \ j = 1, 2, ..., J. \end{aligned} \tag{111}$$

Then we can directly apply the inner approximation method in [34] to problem (111). In each iteration, we keep the objective function t unchanged and approximate all constraints at $\mathbf{X}_0$:

$$\begin{aligned} \underset{\mathbf{X}\in\mathcal{X},t}{\text{maximize}} \quad & t \\ \text{subject to} \quad & \bar{f}_0(\mathbf{X}; \mathbf{X}_0) - t \geq 0 \\ & \bar{f}_j(\mathbf{X}; \mathbf{X}_0) \geq 0, \ j = 1, 2, ..., J \end{aligned} \tag{112}$$

where $\bar{f}_j(\mathbf{X}; \mathbf{X}_0)$ are the corresponding concave lower bound of $f_j(\mathbf{X})$. Note that problem (112) is equivalent to problem (31), which is the subproblem in each iteration of Algorithm 1. Therefore, Algorithm 1 and the proposed algorithm in [34] are equivalent. Based on Theorem 1 in [34], Algorithm 1 converges to the KKT point of problem (5). This completes the proof. ■

Proof of Proposition 2: Recall that problem (39) is

$$R = \underset{\mathbf{X}\in\mathcal{X}}{\text{maximize}} \quad \sum_{i=1}^{I} \big[F_i(\mathbf{X})\big]^{+} \tag{113}$$

where $F_i(\mathbf{X})$ is given by

$$F_i(\mathbf{X}) = \min_{1 \leq j \leq J} f_{ij}(\mathbf{X}). \tag{114}$$

Define the set of nonnegative $F_i(\mathbf{X})$ at the optimal solution of problem (113) as $\mathcal{S} = \{i | F_i(\mathbf{X}_{\text{opt}}) \geq 0, i = 1, 2, ..., I\}$, where $\mathbf{X}_{\text{opt}}$ denotes the optimal solution of (113). Then problem (113) is equivalent to

$$\underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \quad \sum_{i \in \mathcal{S}} F_i(\mathbf{X}). \tag{115}$$

Note that $\mathcal{S}$ can be any non-empty subset of $\{1, 2, ..., I\}$, i.e., $\mathcal{S} \in \{\mathcal{S}_m, m = 1, 2, ..., 2^I - 1\}$, where $\mathcal{S}_m$ is the m-th non-empty subset of $\{1, 2, ..., I\}$. Therefore, we need to solve problem (115) $2^I - 1$ times with different $\mathcal{S}_m$ to obtain R. This completes the proof. ∎

Proof of Proposition 3: We first show that $\{F(\mathbf{X}_n^{\text{opt}})\}$ is monotonically nondecreasing by the following inequalities:

$$F(\mathbf{X}_{n+1}^{\text{opt}}) \geq \bar{F}(\mathbf{X}_{n+1}^{\text{opt}}; \mathbf{X}_n^{\text{opt}}) \geq \bar{F}(\mathbf{X}_n^{\text{opt}}; \mathbf{X}_n^{\text{opt}}) = F(\mathbf{X}_n^{\text{opt}}) \tag{116}$$

where the first inequality holds because $\bar{F}(\cdot; \mathbf{X}_n^{\text{opt}})$ serves as the lower bound of $F(\cdot)$; the second inequality holds because $\mathbf{X}_{n+1}^{\text{opt}}$ is the optimal solution of $\underset{\mathbf{X} \in \mathcal{X}}{\text{maximize}} \, \bar{F}(\mathbf{X}; \mathbf{X}_n^{\text{opt}})$. Furthermore, since the feasible set $\mathcal{X}$ is compact and $F(\cdot)$ is continuous, the sequence $\{F(\mathbf{X}_n^{\text{opt}})\}$ is bounded above. Therefore, the convergence is guaranteed since every bounded monotone sequence has a limit. This completes the proof. ∎

Proof of Proposition 4: The convexity can be shown by reformulating $g(\mathbf{W}; N)$ as

$$\frac{1}{M} \sum_{m=1}^{M} \log \sum_{n=1}^{M} \exp \left[-N \cdot \ln \left(1 + \frac{1}{2} \mathbf{e}_{mn}^{\text{H}} \mathbf{W} \mathbf{e}_{mn} \right) \right]. \tag{117}$$

Since $-\ln \left(1 + 0.5 \cdot \mathbf{e}_{mn}^{\text{H}} \mathbf{W} \mathbf{e}_{mn} \right)$ is a convex function of $\mathbf{W}$ for any vector $\mathbf{e}_{mn}$, and $\log \sum_i \exp(f_i)$ is convex whenever f_i is convex for all i, $g(\mathbf{W}; N)$ is a convex function.

The MNI property can be shown by computing the complex gradient of $g(\mathbf{W}; N)$

$$\nabla_{\mathbf{w}} g(\mathbf{W}, N) = -\sum_{m,n} \alpha_{mn} \mathbf{e}_{mn} \mathbf{e}_{mn}^{\text{H}} \tag{118}$$

where

$$\alpha_{mn} = \frac{N}{2M} \cdot \frac{\left(1 + 0.5 \cdot \mathbf{e}_{mn}^{\text{H}} \mathbf{W} \mathbf{e}_{mn} \right)^{-N-1}}{\sum_{n=1}^{M} \left(1 + 0.5 \cdot \mathbf{e}_{mn}^{\text{H}} \mathbf{W} \mathbf{e}_{mn} \right)^{-N}} > 0. \tag{119}$$

Since $\nabla_{\mathbf{w}} g(\mathbf{W}, N) \preceq \mathbf{0}$, $g(\mathbf{W}, N)$ is a MNI function [23, ch. 3.6.1]. This completes the proof. ∎

REFERENCES

[1] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: SIAM, 2001.

[2] P. D. Tao *et al.*, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Ann. Oper. Res.*, vol. 133, no. 1–4, pp. 23–46, Jan. 2005.

[3] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Comput.*, vol. 15, no. 4, pp. 915–936, Apr. 2003.

[4] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optim. Eng.*, vol. 17, no. 2, pp. 263–287, Jun. 2016.

[5] T. D. Quoc and M. Diehl, "Sequential convex programming methods for solving nonlinear optimization problems with dc constraints," arXiv paper, arXiv:1107.5841, 2011.

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

[7] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 804–808, Jul. 2015.

[8] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, Jul. 2007.

[9] K. Lange and H. Zhou, "MM algorithms for geometric and signomial programming," *Math. Program.*, vol. 143, no. 1/2, pp. 339–356, Feb. 2014.

[10] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Stat.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[11] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[12] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[13] C. Xiao and Y. R. Zheng, "On the mutual information and power allocation for vector Gaussian channels with finite discrete inputs," in *Proc. IEEE Globecom*, 2008, pp. 1–5.

[14] F. Pérez-Cruz, M. R. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[15] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3301–3314, Jul. 2011.

[16] Y. Wu, C. Xiao, Z. Ding, X. Gao, and S. Jin, "Linear precoding for finite-alphabet signaling over MIMOME wiretap channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2599–2612, Jul. 2012.

[17] Y. Wu, M. Wang, C. Xiao, Z. Ding, and X. Gao, "Linear precoding for MIMO broadcast channels with finite-alphabet constraints," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2906–2920, Aug. 2012.

[18] W. Zeng, C. Xiao, J. Lu, and K. B. Letaief, "Globally optimal precoder design with finite-alphabet inputs for cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 1861–1874, Nov. 2012.

[19] Y. Wu, C. Xiao, X. Gao, J. D. Matyjas, and Z. Ding, "Linear precoder design for MIMO interference channels with finite-alphabet signaling," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3766–3780, Sep. 2013.

[20] J. Jin, C. Xiao, M. Tao, and W. Chen, "Linear precoding for fading cognitive multiple access wiretap channel with finite-alphabet inputs," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3059–3070, Apr. 2017.

[21] J. Jin, Y. R. Zheng, W. Chen, and C. Xiao, "Generalized quadratic matrix programming: A unified approach for linear precoder design," in *Proc. IEEE GLOBECOM*, 2016, pp. 1–6.

[22] A. Beck, "Quadratic matrix programming," *SIAM J. Optim.*, vol. 17, no. 4, pp. 1224–1238, Jan. 2007.

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

[24] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

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

[26] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer-Verlag, 1999.

[27] M. Chen, S. C. Liew, Z. Shao, and C. Kai, "Markov approximation for combinatorial network optimization," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6301–6327, Oct. 2013.

[28] 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.

[29] C. Xiao, J. Wu, S.-Y. Leong, Y. R. Zheng, and K. Letaief, "A discrete-time model for triply selective MIMO Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1678–1688, Sep. 2004.

[30] Y. Liang, G. Kramer, H. V. Poor, and S. Shamai, "Compound wiretap channels," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, no. 1, pp. 1–12, Oct. 2009.

[31] W. Zeng, C. Xiao, M. Wang, and J. Lu, "Linear precoding for finite-alphabet inputs over MIMO fading channels with statistical CSI," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3134–3148, Jun. 2012.

[32] E. A. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[33] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[34] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, Jul. 1978.

Juening Jin received the B.S. degree in electronic engineering from Southeast University, Nanjing, China, in 2013. Since September 2013, he has been working toward the Ph.D. degree in the Network Coding and Transmission Laboratory, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. He is currently a Visiting Scholar with Missouri University of Science and Technology, Rolla, MO, USA. His research interests include convex and nonconvex optimization, and physical layer security.

Yahong Rosa Zheng (M'03–SM'07–F'15) received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 1987, and the M.S. degree from Tsinghua University, Beijing, China, in 1989. She received the Ph.D. degree from Carleton University, Ottawa, ON, Canada, in 2002.

From 1989 to 1996, she was an Engineer and technical staff at several companies, including Beijing Peony Electronics Group, Beijing China; Sagem Australasia, Sydney, NSW, Australia; and Polytronics, Markham, ON, Canada. From 2003 to 2005, she was an NSERC Postdoctoral Fellow with the University of Missouri-Columbia. Since fall 2005, he has been a faculty member in the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, USA, where she is now a Professor. Her research interests include array signal processing, wireless communications, and wireless sensor networks. She has published more than 60 journal papers and more than 100 conference papers in these areas. She served as symposium Cochair for several IEEE conferences, including VTC, GLOBECOM, and ICC. She serves as Associate Editor for the IEEE JOURNAL OF OCEANIC ENGINEERING. Previously, she was an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (2009–2016) and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2006–2008). She received an NSF CAREER award in 2009. She is a Distinguished Lecturer of IEEE Vehicular Technology Society since 2015.

Wen Chen (M'03–SM'11) received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 1990 and 1993, respectively, and the Ph.D. degree from the University of Electro-Communications, Tokyo, Japan, in 1999.

From 1999 to 2001, he was a Researcherin the Japan Society for the Promotion of Sciences. In 2001, he joined the University of Alberta, Edmonton, AB, Canada, starting as a Postdoctoral Fellow in the Information Research Laboratory and continuing as a Research Associate in the Department of Electrical and Computer Engineering. Since 2006, he has been a Full Professor in the Department of Electronic Engineering, Shanghai Jiao Tong University (SJTU), Shanghai, China, where he is also the Director of the Institute for Signal Processing and Systems. In 2014–2015, he was the Dean of electronic engineering and automations with Guilin University of Electronic Technology, Guilin, China. Since 2016, he has been the Chairman of SJTU Intelligent Property Management Corporation. He is the author of 73 papers in IEEE Journals and Transactions and more than 100 papers in IEEE Conference Publications. His research interests include network coding, cooperative communications, cognitive radio, and multiple-inputmultiple-output orthogonal-frequency-division multiplexing systems.

Chengshan Xiao (M'99–SM'02–F'10) received the Bachelor's of Science degree in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1987, the Master's of Science degree in electronic engineering from Tsinghua University, Beijing, China, in 1989, and the Ph.D. degree in electrical engineering from the University of Sydney, Sydney, NSW, Australia, in 1997.

He is a Professor in the Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, USA (formerly the University of Missouri - Rolla). He is currently a Program Director at USA National Science Foundation through intergovernmental personnel act assignment. Previously, he was a senior member of scientific staff with Nortel Networks, Ottawa, ON, Canada, a faculty member at Tsinghua University, Beijing, China, the University of Alberta, Edmonton, AB, Canada, and the University of Missouri, Columbia, MO, USA. He also held Visiting Professor positions in Germany and Hong Kong. His research interests include wireless communications, signal processing, and underwater acoustic communications. He is the holder of three U.S. patents. His invented algorithms have been implemented into Nortel's base station radio products after successful technical field trials and network integration.

Dr. Xiao was an elected member of Board of Governors, a member of Fellow Evaluation Committee, Director of Conference Publications, and Distinguished Lecturer of the IEEE Communications Society. He also served as an Editor, an Area Editor, and the Editor-in-Chief of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS–I. He was the Technical Program Chair of the 2010 IEEE International Conference on Communications, Cape Town, South Africa. He served as the founding Chair of the IEEE Wireless Communications Technical Committee. He received several distinguished awards including 2014 Humboldt Research Award, 2014 IEEE Communications Society Joseph LoCicero Award, and 2015 IEEE Wireless Communications Technical Committee Recognition Award. He is a Fellow of Canadian Academy of Engineering.