IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 30, 2016, accepted October 12, 2016, date of publication November 15, 2016, date of current version January 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2016.2628808

Joint Secure AF Relaying and Artificial Noise Optimization: A Penalized Difference-of-Convex Programming Framework

JIAXIN YANG[1], (Student Member, IEEE), QIANG LI[2], (Member, IEEE), YUNLONG CAI[3], (Senior Member, IEEE), YULONG ZOU[4], (Senior Member, IEEE), LAJOS HANZO[5], (Fellow, IEEE), AND BENOIT CHAMPAGNE[1], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 0E9, Canada
[2]School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[3]Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China
[4]School of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[5]School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, U.K.

Corresponding author: L. Hanzo (lh@ecs.soton.ac.uk)

This work of J. Yang and B. Champagne was supported by the Natural Sciences and Engineering Research Council of Canada under Discovery Grant 105533. The work of Q. Li was supported by the National Natural Science Foundation of China under Grant 61401073 and Grant 61531009.

ABSTRACT Owing to the vulnerability of relay-assisted communications, improving wireless security from a physical layer signal processing perspective is attracting increasing interest. Hence, we address the problem of secure transmission in a relay-assisted network, where a pair of legitimate user equipments (UEs) communicate with the aid of a multiple-input multiple output (MIMO) relay in the presence of multiple eavesdroppers (eves). Assuming imperfect knowledge of the eves' channels, we jointly optimize the power of the source UE, the amplify-and-forward relaying matrix, and the covariance of the artificial noise transmitted by the relay, in order to maximize the received signal-to-interference-plus-noise ratio at the destination, while imposing a set of *robust secrecy constraints*. To tackle the resultant non-convex optimization problem with tractable complexity, a new penalized difference-of-convex (DC) algorithm is proposed, which is specifically designed for solving a class of non-convex semidefinite programs. We show how this penalized DC framework can be invoked for solving our robust secure relaying problem with proven convergence. In addition, to benchmark the proposed algorithm, we subsequently propose a semidefinite relaxation-based exhaustive search approach, which yields an upper bound of the secure relaying problem, however, with significantly higher complexity. Our simulation results show that the proposed solution is capable of ensuring the secrecy of the relay-aided transmission and significantly improving the robustness toward the eves' channel uncertainties as compared with the non-robust counterparts. It is also demonstrated the penalized DC-based method advocated yields a performance close to the upper bound.

INDEX TERMS Amplify-and-forward, difference-of-convex, eavesdropping, multiple-input multiple-output, physical layer security, relaying, robust optimization.

I. INTRODUCTION

With the proliferation of smartphones storing more sensitive personal data ranging from social networking to online banking, wireless end-users have become vulnerable targets of hackers. According to a recent report on mobile cyber threats, the number of cyber attacks to mobile users has been dramatically growing, e.g., by nearly 10-fold from August 2013 to March 2014 [1]. Within this context, how to ensure information security is becoming a critical issue for wireless service providers. Although the classic bit-level encryption technique has been deemed to be most effective way of achieving this goal, a recent report by the Washington Post has drawn public attention to the potential security risks of wireless technologies, even when advanced encryption is used.[1] Against this background, physical layer security

[1] In [2], it is reported that two German researchers have demonstrated how to exploit the security flaws in the Signaling System 7 (SS7) to eavesdrop on all incoming and outgoing calls indefinitely from anywhere in the world. They have shown how to decode the messages by requesting each caller's carrier to release a temporary encryption key through the SS7.

is emerging as a promising alternative to complement the encryption and to further enhance the security of wireless networks.

Since Wyner opened this new avenue of security provision by introducing the notion of secrecy capacity [3], researchers have sought to enhance security for a wide range of communication channel models, as discussed in [4]–[6] and the references therein. Recently, physical layer security has attracted increased interest, driven by new techniques such as cooperative relaying, which has found its way into the Long-Term Evolution (LTE) standard. Although the diversity advantages gleaned from user cooperation have been recognized in the context of generic relay-assisted networks [7]–[9], ensuring secrecy in message relaying remains a key issue. Specifically, when additional intermediate nodes assist in forwarding the source messages, the information confidentiality may be more readily compromised, unless the relaying scheme is appropriately designed. It was demonstrated in [10] that relaying is capable of improving the level of security. This seminal work has led to further research endeavors devoted to investigating the secrecy of relay-assisted communications from the physical layer perspective [11]. Following this trend, in this paper emphasis will be on new signal processing techniques conceived for improve wireless relaying security. Below we briefly review related works on this research topic and summarize our main contributions.

A. RELATED WORKS

A wireless relay can adopt either the amplify-and-forward (AF) or the decode-and-forward (DF) strategy for forwarding source messages. For DF relaying, the optimal weights that achieve the maximum secrecy capacity are derived in [12] and [13]. The optimal power allocation scheme betwen the information and jamming signals for the DF relay is derived in [14]. As compared to DF, AF relaying offers its inherent advantages of lower signal processing complexity and latency, and hence will be the focus of our attention. A variety of relaying solutions such as beamforming, cooperative jamming and artificial noise (AN) generation, or a hybrid of the aforementioned options, have been studied in [15]–[26]. For instance, the optimal AF relaying weights maximizing the achievable secrecy rate of a single-antenna relay network are derived in [15], without consideration of the source information leakage to $\mathtt{eves}$. Joint optimization of beamforming, power and jamming signals for single-antenna relay networks is further investigated in [26] with the objective maximizing the secrecy rate. When multiple antennas are employed at both the source and relay, joint transmit precoding and power allocation relying on the generalized singular value decomposition (GSVD) is proposed in [16]. Finally, joint source precoding and multi-antenna AF relaying is investigated in [17] assuming an untrusted relay node.

The contributions [15]–[17] assume perfect knowledge of each $\mathtt{eve}$'s channel state information (ECSI) at the legitimate nodes. In practice, due to the lack of explicit cooperation between the latter and $\mathtt{eves}$, at best an inaccurate estimate of the ECSI may be available. In [27], knowledge of specific distribution of the ECSI errors (e.g., Gaussian) is assumed, and an intercept probability constrained maximum SINR beamforming scheme is proposed for an MIMO relay network. Assuming that the ECSI errors lie in a predefined norm-bounded region, joint relay beamforming and jamming signal design in a single-antenna relay network is developed in [18] and [19] with the objective of maximizing the worst-case secrecy rate. Extension of this approach to a more generalized model where multi-antenna is employed at the relay is considered in [20] and [21], see, also [25] for the scenario of multiple multi-antenna $\mathtt{eves}$. Minimization of the mean square error (MSE) of the received signal at the destination, subject to a set of signal-to-interference-plus-noise (SINR)-based secrecy constraints, is considered in [22]. Using the same uncertainty model, the problem of total relaying power minimization is studied in [23] and [24] by simultaneously guaranteeing a predefined quality-of-service (QoS) level at the destination and a certain secrecy level against eavesdropping. Finally, [21] assumes a more general relay system configuration, where some of the prior works can be viewed as a special case. In this work, a globally optimal solution is obtained resorting to a bi-level optimization framework, where the upper-level problem is tackled by one-dimensional search, while the inner-level problem is solved by semidefinite relaxation (SDR) [28].

B. CONTRIBUTIONS

This paper considers a general wireless communication scenario, where a source ($\mathtt{S}$) transmits its confidential data to a destination ($\mathtt{D}$), assisted by a multi-antenna AF relay ($\mathtt{R}$). Although a similar system model was stuided in [21], the present paper assumes that both phases of the two-hop transmission are overheard by a set of independent $\mathtt{eves}$, which was bypassed by [21]. The power of $\mathtt{S}$, the AF relaying matrix and the covariance matrix of the AN emitted by $\mathtt{R}$ have to be jointly optimized for protecting the message confidentiality. As an alternative to most of the prior contributions [12], [13], [15], [16], [18]–[21], where the main focus has been on the maximization of the (worst-case) secrecy rate when either perfect or imperfect ECSI is available, we investigate the secrecy problem in MIMO relaying network from a practical communication performance perspective. Specifically, assuming that the ECSI errors reside in a predefined spherical region, we aim for maximizing the received SINR at $\mathtt{D}$, subject to power constraints, while satisfying a set of *robust secrecy constraints* at $\mathtt{eves}$. The formulated optimization problem can be represented as a nonlinear nonconvex semidefinite program (SDP) with a bilinear equality constraint due to the joint nature of the optimization variables. Such a class of problems are in general difficult to solve with tractable computational complexity. Towards this end, we propose a new penalized difference-of-convex (DC) algorithmic framework specifically designed for the class of nonliner non-convex SDP with bilinear equality constraints. One of the feature of the proposed penalized DC algorithm is that it eliminates the need for a non-trivial feasible initialization as required by conventional iterative algorithm [29] since finding such an initialization for a non-convex problem is in general a difficult task. We explicitly prove that the solution sequence generated by the algorithm converges to a stationary point of the original problem. We further solve the secrecy constrained relaying problem by the proposed algorithm efficiently. To benchmark our solution approach, we also derive a upper bound for the secrecy constrained relaying problem by relying on the SDR technique along with one-dimensional search algorithm. We show by numerical simulations that our proposed penalized DC algorithm is capable of achieving a performance close to the upper bound at a significantly reduced complexity.

C. ORGANIZATION AND NOTATIONS

The rest of the paper is organized as follows. Section II introduces the relay system model and formulates our secrecy-constrained robust relaying problem. In Section III, we propose a new penalized DC algorithmic framework and characterize its convergence. We then invoke the proposed framework for solving our secure relaying problem in Section IV. In Section V, a benchmarker relying on the SDR and one-dimensional exhaustive search is derived for comparison purpose. The performance of the proposed solution is quantified via numerical simulations in Section VI. Finally, we conclude in Section VII.

Boldface uppercase (lowercase) letters denote matrices (vectors), while normal letters denote scalars; $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote the conjugate, transpose, Hermitian transpose and inverse, respectively; $\|\cdot\|$ represents the Euclidean norm of a vector, while $\|\cdot\|_F$ denotes the Frobenius norm of a matrix; $\mathrm{Tr}(\cdot)$, $\mathrm{vec}(\cdot)$, and $\otimes$ stand for the matrix trace, vectorization and the Kronecker product, respectively; $\mathbb{C}^{M\times M}$ and $\mathbb{H}^M$ denotes the spaces of $(M \times M)$-element matrices having complex entries and $M \times M$ Hermitian matrices, respectively; $\mathrm{Re}\{\cdot\}$ denotes the real part of a complex number.

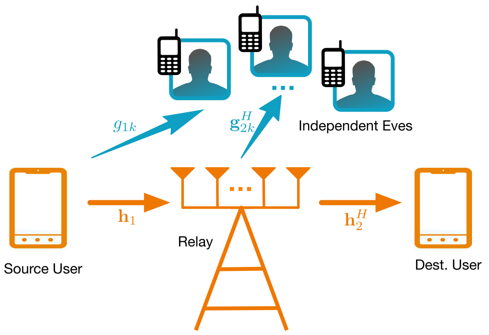


FIGURE 1. MIMO relay network in the presence of multiple single-antenna eves.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider the wireless network as depicted in Fig. 1, where source S communicates with destination D, assisted by a trusted AF relay R operating in a half-duplex mode. The signals transmitted during the S $\to$ R and R $\to$ D hops are overheard by K independent eves, E_k for $k \in \mathcal{K} \triangleq \{1, 2, \cdots, K\}$. We assume that S, D and E_k, $\forall k \in \mathcal{K}$ are single-antenna UEs having limited signal processing capabilities and low power budgets. By contrast, R is equipped with $N_{\mathrm{R}} \geq 2$ antennas. It is assumed that no direct link is available between S–D due to the severe pathloss.

A narrowband flat-fading channel model is considered, where we denote the S–R channel by $\mathbf{h}_1 \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$ and the Hermitian transpose of the R–D channel by $\mathbf{h}_2 \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$. Let s denote the S information symbol, modeled as a zero-mean Gaussian random variable with a power of $\sigma_{\mathrm{S}}^2 \leq P_{\mathrm{S}}$, where P_{S} denotes the S power budget. During the first transmission slot, the signal received at R is given by

$$\mathbf{z} = \mathbf{h}_1 s + \mathbf{n}_{\mathrm{R}}, \tag{1}$$

where $\mathbf{n}_{\mathrm{R}}$ is a zero-mean additive noise vector with covariance of $\sigma_{\mathrm{R}}^2 \mathbf{I}_{N_{\mathrm{R}}}$. Then R applies a linear AF transformation matrix $\mathbf{W} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{R}}}$ to the received signal, and superimposes an AN vector onto the linearly processed signal. Hence, the signal to be forwarded to D is given by

$$\mathbf{r} = \mathbf{W}\mathbf{z} + \mathbf{v} = \mathbf{W}\mathbf{h}_1 s + \mathbf{W}\mathbf{n}_{\mathrm{R}} + \mathbf{v}, \tag{2}$$

where $\mathbf{v}$ denotes the AN vector with zero mean and covariance of $\mathbb{E}\{\mathbf{v}\mathbf{v}^H\} = \boldsymbol{\Psi} \succeq \mathbf{0}$ to be optimized. The relay R has the power constraint of $\sigma_{\mathrm{S}}^2\|\mathbf{W}\mathbf{h}_1\|^2 + \sigma_{\mathrm{R}}^2\|\mathbf{W}\|_F^2 + \mathrm{Tr}(\boldsymbol{\Psi}) \leq P_{\mathrm{R}}$, where P_{R} denotes its power budget. During the second transmission slot, D receives the following signal:

$$y_{\mathrm{D}} = \mathbf{h}_2^H\mathbf{W}\mathbf{h}_1 s + \mathbf{h}_2^H\mathbf{W}\mathbf{n}_{\mathrm{R}} + \mathbf{h}_2^H\mathbf{v} + n_{\mathrm{D}}, \tag{3}$$

where n_{D} is an additive noise with zero mean and a variance of σ_{D}^2.

We adopt, as a metric of transmission reliability, the received SINR at D given by

$$\mathrm{SINR}_{\mathrm{D}} = \frac{\sigma_{\mathrm{S}}^2|\mathbf{h}_2^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{h}_2^H\mathbf{W}\|^2 + \mathbf{h}_2^H\boldsymbol{\Psi}\mathbf{h}_2 + \sigma_{\mathrm{D}}^2}. \tag{4}$$

During the transmission, each E_k is potentially capable of overhearing the signals transmitted both from S and R. Let g_{1k} and $\mathbf{g}_{2k} \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$, respectively, denote the S–E_k channel and the Hermitian transpose of the R–E_k channel. Then the signals observed by E_k from S and R, respectively, are given by

$$y_{\mathrm{E},k}^{\mathrm{S}} = g_{1k}s + n_{\mathrm{E},1k} \tag{5}$$
$$y_{\mathrm{E},k}^{\mathrm{R}} = \mathbf{g}_{2k}^H\mathbf{W}\mathbf{h}_1 s + \mathbf{g}_{2k}^H\mathbf{W}\mathbf{n}_{\mathrm{R}} + \mathbf{g}_{2k}^H\mathbf{v} + n_{\mathrm{E},2k}, \tag{6}$$

where $n_{\mathrm{E},1k}$ and $n_{\mathrm{E},2k}$ are additive noise terms with zero mean and a variance of $\sigma_{\mathrm{E},k}^2$. In our work, it is reasonable to assume that E_k, for $k \in \mathcal{K}$, relies on selection diversity combining of $y_{\mathrm{E},k}^{\mathrm{S}}$ and $y_{\mathrm{E},k}^{\mathrm{R}}$ for the sake of simpler exposition (However, our work can be extended to the case of maximum ratio combining (MRC), see Remark 1 for more justifications.).

On this basis, the mutual information leakage to each $\mathtt{E}_k$ can therefore be expressed as

$$\begin{aligned}&\mathtt{C}_{\mathtt{E},k}(\sigma_{\mathtt{S}}, \mathbf{W}, \mathbf{\Psi})\\ &= \frac{1}{2}\max\left\{\log_2\left(1+\frac{\sigma_{\mathtt{S}}^2|g_{1k}|^2}{\sigma_{\mathtt{E},k}^2}\right),\right.\\ &\qquad \left.\log_2\left(1+\frac{\sigma_{\mathtt{S}}^2|\mathbf{g}_{2k}^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathtt{R}}^2\|\mathbf{g}_{2k}^H\mathbf{W}\|^2+\mathbf{g}_{2k}^H\mathbf{\Psi}\mathbf{g}_{2k}+\sigma_{\mathtt{E},k}^2}\right)\right\}, \quad (7)\end{aligned}$$

where the coefficient $\frac{1}{2}$ is due to the fact that the relay-assisted transmission requires a pair of orthogonal time slots in half-duplex mode.

In practice, due to the lack of explicit cooperation between the legitimate UEs and $\mathtt{eves}$, only imperfect estimates of the ECSI may be available at the legitimate UEs. Like most of the prior contributions in the robust transceiver design literature, we model the unknown ECSI by taking into account the error terms Δg_{1k} and $\Delta \mathbf{g}_{2k}$, yielding:

$$g_{1k} = \hat{g}_{1k} + \Delta g_{1k}, \ \mathbf{g}_{2k} = \hat{\mathbf{g}}_{2k} + \Delta \mathbf{g}_{2k}, \quad (8)$$

where $\hat{g}_{1k}$ and $\hat{\mathbf{g}}_{2k}$ denote the imperfect ECSI estimates, while again, Δg_{1k} and $\Delta \mathbf{g}_{2k}$ represent the corresponding *uncertainties*. Hereby we assume that the ECSI errors lie in some predefined bounded sets, yielding:

$$\mathcal{G}_{1k} \triangleq \left\{\Delta g_{1k} : |\Delta g_{1k}|^2 \le \varepsilon_{1k}\right\} \quad (9)$$

$$\mathcal{G}_{2k} \triangleq \left\{\Delta \mathbf{g}_{2k} : \|\Delta \mathbf{g}_{2k}\|^2 \le \varepsilon_{2k}\right\}, \quad (10)$$

where $\varepsilon_{ik},\ i = 1, 2$ denotes the radius of the uncertainty region. The above bounded error model has been extensively used in robust MIMO transceiver optimization literature to capture the effects of channel estimation errors or quantization errors due to the finite-rate feedback, see, e.g., [30] for more details. The above error model is also applicable in some secure communication scenarios. A notable example is the device-to-device (D2D) discovery and communication defined in 3GPP LTE Rel. 12 [31]. Each UE (including the potential $\mathtt{eves}$) periodically broadcasts its own beacon signals and listens to others using a subset of resources reserved for D2D operations. In this way, each UE is able to discover the presence of other UEs (including potential $\mathtt{eves}$ in its proximity) and subsequently infers an imprecise ECSI estimate based on the channel reciprocity. In this case, the bounded error model can be invoked to quantify the channel estimation errors.

In a practical communication system, $\mathtt{S}$ can operate at a fixed data rate of R_d with specific modulation and coding scheme (MCS), i.e., during a specific scheduling period in LTE. The objective of our secure relaying design is to jointly optimize $\sigma_{\mathtt{S}}$, $\mathbf{W}$ and $\mathbf{\Psi}$, subject to the power constraints, in order to maximize the received SINR at the legitimate end-user $\mathtt{D}$, while satisfying a set of *robust secrecy constraints* at the $\mathtt{eves}$. Mathematically, this problem can be formulated as

$$\max_{\sigma_{\mathtt{S}}, \mathbf{W}, \mathbf{\Psi}} \ \mathtt{SINR}_{\mathtt{D}} \quad (11a)$$

$$\text{s.t. } \mathtt{C}_{\mathtt{E},k}(\sigma_{\mathtt{S}}, \mathbf{W}, \mathbf{\Psi}; \Delta g_{1k}, \Delta \mathbf{g}_{2k}) \le \kappa R_d,$$

$$\forall \Delta g_{1k} \in \mathcal{G}_{1k}, \Delta \mathbf{g}_{2k} \in \mathcal{G}_{2k}, k \in \mathcal{K} \quad (11b)$$

$$\sigma_{\mathtt{S}}^2\|\mathbf{W}\mathbf{h}_1\|^2 + \sigma_{\mathtt{R}}^2\|\mathbf{W}\|_F^2 + \mathrm{Tr}(\mathbf{\Psi}) \le P_{\mathtt{R}} \quad (11c)$$

$$\sigma_{\mathtt{S}}^2 \le P_{\mathtt{S}}, \ \mathbf{\Psi} \succeq \mathbf{0}. \quad (11d)$$

In the above formulation, (11b) denotes the so-called *robust secrecy constraints*, which aims to guarantee the secrecy for all possible realizations of the ECSI errors Δg_{1k} and $\Delta \mathbf{g}_{2k}$ within uncertainty regions as defined in (9) and (10), respectively. The parameter κ is used to introduce more flexibility in controlling the security level of the communication. Before leaving this section, two important remarks are presented:

Remark 1 (On the assumption of $\mathtt{eves}$*' receive combining):* It is worth pointing out that in contrast to prior contributions, hereby we assume information leakage during both the two-hop relay-assisted transmission. This more general assumption grants the $\mathtt{eves}$ the opportunities of enhancing their quality of reception via diversity combining. Two popular diversity combing schemes are available, namely, SC and MRC. The implement of MRC requires an accurate estimate of the phases of the received signals during the two stages of relay-assisted transmission. When channel estimation errors are in general invoked, the performance of MRC would significantly deteriorate. Additionally, to coherently combine the signals from the two-stage transmission, $\mathtt{eves}$' clocks need to be perfectly synchronized to that of the legitimate network, which is quite challenging if the $\mathtt{eves}$ are not part of the legitimate network. It is observed in [32] that the MRC with two branches only yields marginal performance gain over the SC, however, at the expense of higher complexity. Hence, to bypass the aforementioned requirements, it is reasonable to assume that $\mathtt{eves}$ adopt the SC, also for the sake of lower hardware complexity. However, to better appreciate the generality of our proposed algorithm, in Remark 4 of Section IV, we will elaborate on how the proposed algorithm can be applied to solve the secure relaying problem when the MRC is employed by $\mathtt{eves}$. □

Remark 2 (On the problem formulation of (11d)*):* In literature, another popular approach for improving the transmission secrecy is to maximize the secrecy capacity of the relay-assisted network from the perspective of information theory. The latter in general relies on the underlying assumption that there exists a capacity-achieving coding scheme based on non-constructive random coding theorem. Such design approach is therefore useful as a benchmark from system design viewpoint. In practical communication systems whereby specific MCSs are used, e.g., 3GPP LTE-Advanced, it is better to consider a physical layer design approach, which can be readily incorporated into on-going standards. The proposed design approach well suits several use cases in LTE-Advanced such as the D2D broadcast scenarios. Specifically, by enforcing the mutual information leakage $\mathtt{C}_{\mathtt{E},k}$ to fall below the data rate of the legitimate UE, i.e., $\mathtt{C}_{\mathtt{E},k} < \kappa R_d$, $\mathtt{eves}$ are impossible to perfectly decode the confidential messages from the legitimate UEs. □

III. THEORY: PENALIZED DC ALGORITHMIC FRAMEWORK

In this section, we propose a new penalized DC algorithmic framework, which aims to solve a class of nonlinear nonconvex SDPs. Following some preliminary, we first present the framework, which can be considered as an evolutionary variant of the conventional DC framework [33]. However, the results of convergence analysis in the literature of conventional DC algorithm is not directly applicable to the proposed framework. Hence as a further contribution, we explicitly state the convergence properties of this new algorithm.

A. PRELIMINARY

We first provide some definitions which will be used throughout the subsequent derivations of algorithm.

Definition 1 (Positive Semi-Definite (PSD)-Convex Mapping): A matrix-valued mapping $\mathcal{F}(\cdot) : \mathbb{C}^n \to \mathbb{H}^p$ is called *PSD*-convex on a convex subset $\Omega \subseteq \mathbb{C}^n$, if for all $\mathbf{x}, \mathbf{y} \in \Omega$ and θ with $0 \le \theta \le 1$, we have

$$\mathcal{F}(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \preceq \theta \mathcal{F}(\mathbf{x}) + (1-\theta)\mathcal{F}(\mathbf{y}). \tag{12}$$

The PSD-convex mapping is a generalization of a convex function by noting that any convex function with $f(\cdot) : \mathbb{C}^n \to \mathbb{R}$ is PSD-convex in conjunction with $p = 1$. The derivative of a matrix-valued mapping $\mathcal{F}(\cdot)$ at a point $\mathbf{x}$ is defined as a linear mapping $\mathcal{DF} : \mathbb{C}^n \to \mathbb{C}^{p\times p}$ given by

Definition 2 (Directional Derivative of Matrix-Valued Mapping): The directional derivative of a matrix-valued mapping $\mathcal{F}$ at $\mathbf{x}$ is a linear mapping $\mathcal{DF} : \mathbb{C}^n \to \mathbb{C}^{p\times p}$, which is defined by

$$\mathcal{DF}\mathbf{h} = \sum_{i=1}^{n} h_i \frac{\partial \mathcal{F}}{\partial x_i}(\mathbf{x}), \ \forall \mathbf{h} \in \mathbb{C}^n. \tag{13}$$

For a given convex subset $\Omega \subseteq \mathbb{C}^n$, the matrix-valued mapping $\mathcal{F}(\cdot)$ is said to be differentiable on Ω if its directional derivative $\mathcal{DF}$ exits at every $\mathbf{x} \in \Omega$. For ease of discussion, we assume that all the functions and matrix-valued mappings are twice differentiable on their corresponding domains throughout the paper.

The first-order condition for a PSD-convex mapping is given in the following proposition:

Proposition 1 (First-Order Condition): A mapping $\mathcal{F}$ is PSD-convex if and only if for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, the following inequality holds

$$\mathcal{F}(\mathbf{y}) \succeq \mathcal{F}(\mathbf{x}) + \mathcal{DF}(\mathbf{x})(\mathbf{y}-\mathbf{x}). \tag{14}$$

Now we can proceed to the definition of a PSD DC mapping.

Definition 3 (PSD DC Mapping): A matrix-valued mapping $\mathcal{H}(\cdot)$ is called a PSD DC mapping if $\mathcal{H}$ can be represented as a difference of two PSD-convex mappings, i.e.,

$$\mathcal{H}(\mathbf{x}) = \mathcal{F}(\mathbf{x}) - \mathcal{G}(\mathbf{x}), \tag{15}$$

where $\mathcal{F}(\cdot)$ and $\mathcal{G}(\cdot)$ are PSD-convex mappings. Note that the concept of the PSD DC mapping generalizes the conventional scalar-valued DC scalar-valued function, i.e., $h(\mathbf{x}) = f(\mathbf{x}) - g(\mathbf{x})$.

B. OPTIMIZATION OF A PSD DC PROGRAM WITH BILINEAR MATRIX EQUALITY CONSTRAINT

To simplify the exposition, in this subsection let us use matrix $\mathbf{X} \in \mathbb{C}^{m\times n}$ as an optimization variable instead of using $\mathbf{x} \in \mathbb{C}^n$. The reason is that in the problem formulation of our interest, there exists a bilinear matrix equality constraint, as will seen below. However, it should be pointed out that any matrix variable $\mathbf{X} \in \mathbb{C}^{m\times n}$ can be equivalently expressed in the vector form, i.e., $\mathbf{x} \in \mathbb{C}^{mn\times 1}$ via $\mathbf{x} = \mathrm{vec}(\mathbf{X})$. Since the vectorization is a linear operation, the aforementioned PSD-convexity is preserved under linear operation.

We are interested in solving the following problem:

$$\min_{\mathbf{X}} \ f_0(\mathbf{X}) - g_0(\mathbf{X}) \tag{16a}$$
$$\text{s.t.} \ \mathcal{F}_i(\mathbf{X}) - \mathcal{G}_i(\mathbf{X}) \preceq \mathbf{0}, \ i = 1, 2, \cdots, I-1 \tag{16b}$$
$$\mathbf{X}_2 = \mathbf{X}_0\mathbf{X}_1 \tag{16c}$$
$$\mathbf{X} \in \Omega, \tag{16d}$$

where the optimization variable $\mathbf{X}$ is defined as $\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_{N-1})$, $\Omega \subseteq \mathbb{C}^n$ is a non-empty, closed convex subset, $f_0(\cdot)$, $g_0(\cdot)$ are convex functions on Ω, and $\mathcal{F}_i(\cdot)$, $\mathcal{G}_i(\cdot)$ are PSD-convex mappings on Ω. For the ease of presentation, we use (16c) to represent that some of the optimization variables are nonlinearly coupled in the bilinear form. However, it can be conveniently extended to the case of $\mathbf{X}_i = \mathbf{X}_j\mathbf{X}_k$ for $i, j, k \in \{0, 1, \cdots, N-1\}$. Clearly, if the matrix equality constraint (16c) is absent, then (16) becomes a so-called PSD DC program.

Next, we rely on the following lemma to show that (16) can be equivalently rewritten as a PSD DC program.

Lemma 1 (Lemma 1, [34]): Given $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{X}_2$ of appropriate dimensions, which satisfy the following relation:

$$\mathbf{X}_2 = \mathbf{X}_0\mathbf{X}_1, \tag{17}$$

then the above matrix equality is equivalent to the following two constraints:

$$\begin{bmatrix} \mathbf{Y}_1 & \mathbf{X}_2 & \mathbf{X}_0 \\ \mathbf{X}_2^H & \mathbf{Y}_2 & \mathbf{X}_1^H \\ \mathbf{X}_0^H & \mathbf{X}_1 & \mathbf{I} \end{bmatrix} \succeq \mathbf{0} \tag{18}$$

$$\mathrm{Tr}\,(\mathbf{Y}_1) - \mathrm{Tr}\left(\mathbf{X}_0\mathbf{X}_0^H\right) \le 0 \tag{19}$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are auxiliary matrix variables with appropriate dimensions. □

It is observed that (18) is a linear matrix inequality (LMI) constraint and (19) is a DC constraint. Therefore, we can conveniently embed (18) into the convex subset Ω and its convexity remains unaffected. Additionally, since the DC function in (19) is a special case of the PSD DC mapping with $p = 1$, we can incorporate (19) into (16b), and re-express (16) as a standard PSD DC program, which is defined as follows:

Definition 4 (PSD DC Program): A *PSD* DC program assumes the form of

$$\min_{\mathbf{x}} \ \varphi(\mathbf{x}) \triangleq f_0(\mathbf{x}) - g_0(\mathbf{x}) \tag{20a}$$

$$\text{s.t. } \boldsymbol{\mathcal{F}}_i(\mathbf{x}) - \boldsymbol{\mathcal{G}}_i(\mathbf{x}) \preceq \mathbf{0}, \ \ i \in \mathcal{I} \triangleq \{1, 2, \cdots, I\} \tag{20b}$$

$$\mathbf{x} \in \Omega, \tag{20c}$$

where $\mathbf{x}$ collectively denotes all the optimization variables and auxiliary variables with appropriate linear transformation, i.e., $\mathbf{x} \triangleq (\text{vec}(\mathbf{X}), \text{vec}(\mathbf{Y}_1), \text{vec}(\mathbf{Y}_2))$. The above PSD DC program represents a generalization of the conventional DC program [29], where the DC inequality constraint, e.g., $f_i(\mathbf{x}) - g_i(\mathbf{x}) \leq 0$ is now extended to the *generalized inequality* $\preceq$ on the PSD cone. If the convex subset Ω is a polyhedral (which is true for most MIMO-aided transceiver optimization problems), the formulation in (20) can properly represent several classes of optimization problems:

- If at least one of f_0, g_0, $\boldsymbol{\mathcal{F}}_i$ and $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{I}$ is nonlinear, then (20) is a nonlinear SDP;
- If g_0 and $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{I}$ are linear, then (20) subsequently becomes a convex nonlinear SDP
- If at least one of g_0 and $\boldsymbol{\mathcal{G}}_i$ for $i \in \mathcal{I}$ are nonlinear, (20) represents a general nonlinear non-convex SDP.

C. ISSUES WITH THE CONVENTIONAL DC ALGORITHM

Since (20) can be considered as a direct extension of a conventional DC program involving only scalar-valued functions, a natural question arises as to whether the conventional DC algorithm developed in [29] is applicable to solving (20)? Following the line of [29], an iterative algorithm can be developed for (20), where the key ingredient is to find a local linear approximation of the non-convex parts of the objective function (20a) and the PSD DC constraints (20b), i.e., $-g_0(\cdot)$ and $-\boldsymbol{\mathcal{G}}_i(\cdot)$, around the solution $\mathbf{x}^{(n-1)}$ obtained in the previous iteration, such that the resultant sub-problem becomes a convex SDP. The original non-convex problem can then be iteratively solved by a sequence of these "convexified" SDPs. Assuming that $\mathbf{x}^{(n)}$ is a solution obtained at the n^{th} iteration, the linearized sub-problem is then given by

$$\min_{\mathbf{x}} f_0(\mathbf{x}) - g_0(\mathbf{x}^{(n)}) - \nabla g_0^T(\mathbf{x}^{(n)})(\mathbf{x} - \mathbf{x}^{(n)}) \tag{21a}$$

$$\text{s.t. } \boldsymbol{\mathcal{F}}_i(\mathbf{x}) - \boldsymbol{\mathcal{G}}_i(\mathbf{x}^{(n)}) - \mathcal{D}\boldsymbol{\mathcal{G}}_i(\mathbf{x}^{(n)})(\mathbf{x} - \mathbf{x}^{(n)}) \preceq \mathbf{0}, \ \ i \in \mathcal{I} \tag{21b}$$

$$\mathbf{x} \in \Omega. \tag{21c}$$

Since f_0 and $\boldsymbol{\mathcal{F}}_i$ are convex function/mapping in $\mathbf{x}$ and the remaining terms are linear in $\mathbf{x}$, the above problem is a convex (nonlinear) SDP. The iterative algorithm therefore generates a sequence of intermediate solutions $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$. Before proceeding to analyze the feasibility of $\{\mathbf{x}^{(n)}\}$, we first define the feasible set of the original PSD DC program in (20) as

$$\mathcal{D} \triangleq \{\mathbf{x} \in \Omega : \boldsymbol{\mathcal{F}}_i(\mathbf{x}) - \boldsymbol{\mathcal{G}}_i(\mathbf{x}) \preceq \mathbf{0}, \ \ i \in \mathcal{I}\}, \tag{22}$$

and the relative interior of $\mathcal{D}$ as

$$\text{ri}(\mathcal{D}) \triangleq \{\mathbf{x} \in \text{ri}(\Omega) : \boldsymbol{\mathcal{F}}_i(\mathbf{x}) - \boldsymbol{\mathcal{G}}_i(\mathbf{x}) \prec \mathbf{0}, \ \ i \in \mathcal{I}\}. \tag{23}$$

In order to guarantee that the obtained solution sequence $\{\mathbf{x}^{(n)}\}$ lies in the feasible set $\mathcal{D}$, a strictly feasible initialization, i.e., $\mathbf{x}^{(0)} \in \text{ri}(\Omega)$ is required by the conventional DC algorithm.[2] Hence, the following requirements are necessary for the conventional DC algorithm:

Requirement 1 A strictly feasible initialization $\mathbf{x}^{(0)} \in \text{ri}(\mathcal{D})$ is required by the conventional DC algorithm.
Subsequently, it is straightforward to have

Requirement 2 The relative interior of the feasible set is nonempnty, i.e., $\text{ri}(\mathcal{D}) \neq \emptyset$.

We now explain the practical difficulties in satisfying the above requirements. As mentioned earlier, since $\mathcal{D}$ is a non-convex set, finding a strictly feasible initialization within a non-convex set corresponds to the following non-convex feasibility search problem

$$\text{Find } \mathbf{x} \quad \text{s.t. } \mathbf{x} \in \mathcal{D}, \tag{24}$$

which in principle is not a simple task. In fact, solving the above feasibility search problem would require the same amount of computational efforts as solving the original PSD DC program (20). Otherwise, if the algorithm starts with an infeasible point, then it can lead to further infeasibility problems during the successive iterations.

Additionally, the following claim also prevents the direct application of the conventional DC algorithm to (20).

Claim 1 The relative interior of the feasible set of (20) is empty, i.e., $\text{ri}(\mathcal{D}) = \emptyset$.

Proof: We show by contradiction. Recall that a strictly feasible solution to (20) has to satisfy

$$\text{Tr}(\mathbf{Y}_1) - \text{Tr}(\mathbf{X}_0\mathbf{X}_0^H) < 0. \tag{25}$$

By applying the Schur complement to (18), we have

$$\begin{aligned} &\begin{bmatrix} \mathbf{Y}_1 & \mathbf{X}_2 \\ \mathbf{X}_2^H & \mathbf{Y}_2 \end{bmatrix} - \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1^H \end{bmatrix} \begin{bmatrix} \mathbf{X}_0^H & \mathbf{X}_1 \end{bmatrix} \succeq \mathbf{0} \\ \Longleftrightarrow &\begin{bmatrix} \mathbf{Y}_1 & \mathbf{X}_2 \\ \mathbf{X}_2^H & \mathbf{Y}_2 \end{bmatrix} - \begin{bmatrix} \mathbf{X}_0\mathbf{X}_0^H & \mathbf{X}_0\mathbf{X}_1 \\ \mathbf{X}_1^H\mathbf{X}_0^H & \mathbf{X}_1^H\mathbf{X}_1 \end{bmatrix} \succeq \mathbf{0} \\ \Longrightarrow &\ \mathbf{Y}_1 \succeq \mathbf{X}_0\mathbf{X}_0^H, \end{aligned} \tag{26}$$

which obviously contradicts (25). Therefore, we must have

$$\text{Tr}(\mathbf{Y}_1) - \text{Tr}(\mathbf{X}_0\mathbf{X}_0^H) = 0, \tag{27}$$

which implies that $\text{ri}(\mathcal{D}) = \emptyset$. ■

Based on the above analysis, it is known both requirements of the conventional DC algorithm cannot be satisfied. Motivated by the latter, we shall propose a new approach where the concept of penalized DC algorithm is developed for the considered PSD DC program. The proposed penalized DC algorithm, which can be considered as an evolutionary variant of the conventional DC algorithm, can solve a wider range of PSD DC programs. In particular, it eliminates the requirements of a non-trivial initialization and of a feasible set with non-empty relative interior.

[2] This is due to the fact that the first-order Taylor series expansion of the concave function $-\boldsymbol{\mathcal{G}}_i(\cdot)$ is its upper bound, such that we have $\boldsymbol{\mathcal{F}}_i(\mathbf{x}) - \boldsymbol{\mathcal{G}}_i(\mathbf{x}) \preceq \boldsymbol{\mathcal{F}}_i(\mathbf{x}) - \boldsymbol{\mathcal{G}}_i(\mathbf{x}^{(n)}) - \mathcal{D}\boldsymbol{\mathcal{G}}_i(\mathbf{x}^{(n)})(\mathbf{x} - \mathbf{x}^{(n)}) \preceq \mathbf{0}$ for all $\mathbf{x} \in \Omega$.

D. PENALIZED PSD DC ALGORITHMIC FRAMEWORK

Instead of solving (21), hereby we introduce a set of matrix auxiliary variables $\{\mathbf{S}_i\}_{i=1}^I$ and penalize (21a) with a linear regularization term, i.e.,

$$\begin{aligned}\min_{\mathbf{x},\mathbf{S}}\ & \hat{\varphi}^{(n)}(\mathbf{x},\mathbf{S};\mathbf{x}^{(n)}) \\ & \triangleq f_0(\mathbf{x}) - g_0(\mathbf{x}^{(n)}) \\ & \quad - \nabla g_0^T(\mathbf{x}^{(n)})(\mathbf{x}-\mathbf{x}^{(n)}) + \tau^{(n)}\sum_{i=1}^{I}\mathrm{Tr}(\mathbf{S}_i) \end{aligned} \tag{28a}$$

$$\text{s.t. } \mathcal{F}_i(\mathbf{x}) - \mathcal{G}_i(\mathbf{x}^{(n)}) - \mathcal{DG}_i(\mathbf{x}^{(n)})(\mathbf{x}-\mathbf{x}^{(n)}) \preceq \mathbf{S}_i \tag{28b}$$

$$\mathbf{S}_i \succeq \mathbf{0},\ i \in \mathcal{I} \tag{28c}$$

$$\mathbf{x} \in \Omega, \tag{28d}$$

where $\tau^{(n)} \geq 0$ denotes the weight associated with the penalty term at the n^{th} iteration and $\mathbf{S}$ collectively denotes $\mathbf{S} \triangleq (\mathbf{S}_1, \cdots, \mathbf{S}_I)$. The auxiliary variable $\mathbf{S}_i \in \mathbb{H}^{p_i}$ can be viewed as an abstract measure of the extent to which the i^{th} constraint in (21b) is violated. Specifically, $\mathrm{Tr}(\mathbf{S}_i) = 0$ reveals that the i^{th} constraint is satisfied while $\mathrm{Tr}(\mathbf{S}_i) > 0$ indicates the opposite. Therefore, a feasible solution $\mathbf{x} \in \Omega$ is found if

$$\sum_{i=1}^{I}\mathrm{Tr}(\mathbf{S}_i) = 0. \tag{29}$$

With the introduction of the penalized sub-problem (28), we now develop an iterative procedure for solving the PSD DC program (20). The rationale of the proposed penalized DC algorithm is that it starts with an arbitrary point within the convex subset Ω, i.e., $\mathbf{x}^{(0)} \in \Omega$, as opposed to $\mathbf{x}^{(0)} \in \mathrm{ri}(\mathcal{D})$ (hence possibly infeasible), and a small penalty τ such that it facilitates a fast descent of the objective function at the beginning while the constraints are temporarily allowed to be violated, i.e, $\mathbf{S}_i \succ \mathbf{0}$. As iterations evolve, the value of τ gradually increases according to some designed rule in order to enforce the solution to be closer to and finally lie in the feasible region $\mathcal{D}$.

The penalized DC algorithm, which iteratively solves a sequence of sub-problems (28) with a specifically designed updating rule of τ is then described as Algorithm 1.

We now discuss a few important implementation aspects of Algorithm 1.

1) Initialization: Instead of finding an initialization within the relative interior of a non-convex feasible set [c.f. (23)], i.e., $\mathbf{x}^{(0)} \in \mathrm{ri}(\mathcal{D})$, Algorithm 1 can now be initialized with a point $\mathbf{x}^{(0)} \in \Omega$, which corresponds to a more computationally efficient *convex* feasible search problem. For implementation, one may rely on the general-purpose optimization solvers to find $\mathbf{x}^{(0)}$. More importantly, in many practical problems, $\mathbf{x}^{(0)}$ can be easily found by exploiting the specific structure of the convex subset Ω in that problem, (see the considered secure relaying design problem in Section IV).

2) Termination Criterion: In practical implementation, Algorithm 1 needs to be terminated within a maximum of number iterations. Thus, a reasonable termination criterion is that the successive difference in the solution becomes small, i.e., $\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\| \leq \delta$ and $\mathbf{x}^{(n)}$ is (nearly) feasible, i.e., $\sum_{i=1}^{I}\mathrm{Tr}(\mathbf{S}_i) \approx 0$. If the criterion cannot be satisfied within a maximum number of iterations, we claim that the algorithm fails to find a feasible solution given a limited time frame.

Algorithm 1 Penalized DC Algorithm

```
Intialization: An initial point $\mathbf{x}^{(0)} \in \Omega$, $\tau^{(0)} > 0$, $\delta_1 > 0$
  and $\delta_2 > 0$. Set $n = 0$.
  repeat
    1. Convexify: Compute the first-order approximates
       $g_0(\mathbf{x}) \approx g_0\left(\mathbf{x}^{(n)}\right) + \nabla g_0^T\left(\mathbf{x}^{(n)}\right)\left(\mathbf{x}-\mathbf{x}^{(n)}\right)$
       $\mathcal{G}_i(\mathbf{x}) \approx \mathcal{G}_i\left(\mathbf{x}^{(n)}\right) + \mathcal{DG}_i\left(\mathbf{x}^{(n)}\right)\left(\mathbf{x}-\mathbf{x}^{(n)}\right)$
    2. Solve: Compute $\mathbf{x}^{(n+1)}$ by solving (28)
    3. Update $\tau$: Obtain the dual variable $\mathbf{\Phi}_i^{(n+1)}$
  associated with (28b) and set
       $\tau^{(n+1)} = \begin{cases} \tau^{(n)} & \text{if } \tau^{(n)} \geq r^{(n)} \\ \tau^{(n)} + \delta_2 & \text{if } \tau^{(n)} < r^{(n)} \end{cases}$   (30)
  where
       $r^{(n)} \triangleq \min\left\{\|\mathbf{x}^{(n+1)} - \mathbf{x}^{n}\|^{-1}, \lambda_{\max}\left[\sum_{i=1}^{I}\mathbf{\Phi}_i^{(n+1)}\right] + \delta_1\right\}$
    4. Update iteration: $n \leftarrow n + 1$
  until Termination criterion is satisfied or a maximum
  number of iterations are reached
Output: The optimized $\mathbf{x}^*$.
```

3) On the Updating Rule (30)*:* The updating rule of τ in (30) is motivated by the theory of exact penalty function methods for nonlinear optimization problem [35], [36]. The theory suggests that if the penalty τ is larger than all the dual variables $\{\mathbf{\Phi}_i\}$ associated with (28b) (in our case, it is in the form of PSD ordering), i.e., $\tau\mathbf{I} \succeq \mathbf{\Phi}_i$ for all i, then (28) and (21) become equivalent. Also from the definition of $r^{(n)}$ below (30), we see that the unboundness of $\{\tau^{(n)}\}$ leads to the unboundness of $\{\mathbf{\Phi}_i^{(n)}\}$ and $\|\mathbf{x}^{(n+1)} - \mathbf{x}^{n}\| \to 0$. This key property will be exploited later in proving the convergence of Algorithm 1.

4) Solving the Convex Sub-Problem (28)*:* As mentioned earlier, (28) is a general nonlinear convex SDP, which can be solved by a general interior-point method. To our best knowledge, the external solvers supporting a general *nonlinear* SDP is still limited, i.e., some widely-used solvers such as `SeDuMi` and `MOSEK` do not support nonlinear SDPs at current stage while `PENLAB` is the only public nonlinear SDP solver. However, many MIMO transceiver optimization problems exhibit some common structures. Specifically:

1) The convex subset Ω can be represented by a finite number of LMIS, i.e.,

$$\Omega \triangleq \{\mathbf{x} : \mathcal{A}_l(\mathbf{x}) + \mathbf{C}_l \succeq \mathbf{0}, l = 1, \cdots, L\}, \tag{31}$$

where $\mathcal{A}_l(\mathbf{x})$ is a linear mapping of $\mathbf{x}$.

2) The mappings $\mathcal{F}_i(\mathbf{x})$ for $i \in \mathcal{I}$ are so-called Schur PSD-convex mappings, which assumes the form of

$$\mathcal{F}_i(\mathbf{x}) \triangleq \mathcal{S}_i(\mathbf{x})\mathcal{R}_i^{-1}(\mathbf{x})\mathcal{S}_i^H(\mathbf{x}) - \mathcal{Q}_i(\mathbf{x}), \tag{32}$$

where $\mathcal{R}_i(\mathbf{x}) = \mathcal{R}_i^H(\mathbf{x})$ and $\mathcal{Q}_i(\mathbf{x}) = \mathcal{Q}_i^H(\mathbf{x})$ are linear mappings of $\mathbf{x}$ and $\mathcal{R}_i(\mathbf{x}) \succ \mathbf{0}$;

3) The function $f_0(\mathbf{x})$ in the objective function is quadratic in $\mathbf{x}$:

$$f_0(\mathbf{x}) = \mathbf{x}^H\mathbf{B}\mathbf{x} + 2\,\mathrm{Re}\{\mathbf{b}^H\mathbf{x}\} + c. \tag{33}$$

Below we show the sub-problem (20) with the above structure can be equivalently transformed into a standard SDP, which can be efficiently solved by state-of-the-art optimization tools. The transformation simply invokes the Schur complement and the introduction of auxiliary variables. In this case, one can transform (28) into a standard SDP:

$$\min_{\substack{\mathbf{x},\mathbf{S},\\ t,\{\mathbf{T}_i\}}} \; t - \nabla g_0^T\mathbf{x}^{(n)})\mathbf{x} + \tau^{(n)}\sum_{i=1}^{I}\mathrm{Tr}(\mathbf{S}_i) \tag{34a}$$

$$\text{s.t. } \mathbf{T}_i - \mathcal{D}\mathcal{G}_i(\mathbf{x}^{(n)})\mathbf{x} - \mathbf{S}_i \preceq \mathbf{0} \tag{34b}$$

$$\mathcal{A}_l(\mathbf{x}) + \mathbf{C}_l \succeq \mathbf{0},\; l = 1, \cdots, L \tag{34c}$$

$$\begin{bmatrix} \mathcal{Q}_i(\mathbf{x}) + \mathbf{T}_i & \mathcal{S}_i(\mathbf{x}) \\ \mathcal{S}_i^H(\mathbf{x}) & \mathcal{R}_i(\mathbf{x}) \end{bmatrix} \succeq \mathbf{0},\; i \in \mathcal{I} \tag{34d}$$

$$\begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}^H \begin{bmatrix} \mathbf{B} & \mathbf{b} \\ \mathbf{b}^H & c - t \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \preceq \mathbf{0}, \tag{34e}$$

where t and $\{\mathbf{T}_i\}$ are auxiliary variables. The above problem now is in the form of a standard SDP.

E. CONVERGENCE ANALYSIS OF THE PENALIZED DC ALGORITHM

Since Algorithm 1 is designed to start with a possibly infeasible initialization, the iterative procedure may admit an infeasible final solution to the original PSD DC program (20). Therefore, two important aspects regarding the convergence of Algorithm 1 need to be examined:

1) whether the solution generated by Algorithm 1 is feasible to the PSD DC program (20)?
2) whether the convergence properties of conventional DC algorithm still hold for the penalized DC algorithm?

In this subsection, the convergence properties of Algorithm 1 are analytically established. Let $\bar{\mathbf{x}}$ be a point within the convex subset Ω, i.e., $\bar{\mathbf{x}} \in \Omega$. The PSD DC constraint (20b) at $\bar{\mathbf{x}}$ is called *inactive* if the strict inequality holds, that is, $\mathcal{F}_i(\bar{\mathbf{x}}) - \mathcal{G}_i(\bar{\mathbf{x}}) \prec \mathbf{0}$. Otherwise, the PSD DC constraint is called *active*, i.e., $\mathcal{F}_i(\bar{\mathbf{x}}) - \mathcal{G}_i(\bar{\mathbf{x}}) \not\prec \mathbf{0}$. Let us denote the set of *active* constraints at $\bar{\mathbf{x}}$ by

$$\mathcal{U}(\bar{\mathbf{x}}) \triangleq \left\{ i \in \mathcal{I} \middle| \mathcal{F}_i(\bar{\mathbf{x}}) - \mathcal{G}_i(\bar{\mathbf{x}}) \not\prec \mathbf{0} \right\}. \tag{35}$$

We call a vector $\mathbf{h} \in \mathtt{cone}(\Omega - \bar{\mathbf{x}})$ a *feasible direction* to (20) at $\bar{\mathbf{x}}$ if we have

$$\left(\mathcal{D}\mathcal{F}_i(\bar{\mathbf{x}}) - \mathcal{D}\mathcal{G}_i(\bar{\mathbf{x}})\right)\mathbf{h} \prec \mathbf{0},\; \forall i \in \mathcal{U}(\bar{\mathbf{x}}). \tag{36}$$

We now make our first assumption, which is called the *extended Mangasarian-Fromovitz constraint qualification* (MFCQ) [37]:

Assumption 1 For any $\bar{\mathbf{x}} \in \Omega$, there exists a *feasible direction* $\mathbf{h} \in \mathtt{cone}(\Omega - \bar{\mathbf{x}})$ to (20).

The extended MFCQ is a quite common constraint qualification in nonlinear optimization theory such that it guarantees the KKT necessary conditions to hold at a local point.[3] A geometric interpretation of the extended MFCQ can be described as follows. The gradients of the active inequality constraints (recall that $\mathcal{F}_i(\bar{\mathbf{x}}) - \mathcal{G}_i(\bar{\mathbf{x}}) \not\prec \mathbf{0}$) at $\bar{\mathbf{x}}$ form a pointed cone, and there exists a feasible direction in this cone that is tangent to the surface formed by active inequality constraints.

In addition, we also make the following common assumptions:

Assumption 2 Ω is bounded and the objective function $\varphi(\mathbf{x}) = f_0(\mathbf{x}) - g_0(\mathbf{x})$ is bounded from below on Ω.

Assumptions 2 is a mild assumption from practical perspective. In Assumption 2, Ω is bounded due to the power constraints imposed in the design problem, whilst the objective function is usually a performance metric such as the SINR or MSE, which is lower-bounded by zero.

Before formally stating the convergence theorem, we first present the following lemma, which shows that $\Delta\mathbf{x}^{(n)} \triangleq \mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}$ is a descent direction of the PSD DC program (20). The latter is a key property in proving the convergence of Algorithm 1.

Lemma 2 Let us denote the penalized objective function by $\hat{\varphi}^{(n)}(\mathbf{x}, \mathbf{S}) \triangleq f_0(\mathbf{x}) - g_0(\mathbf{x}) + \sum_{i=1}^{I}\tau^{(n)}\,\mathrm{Tr}(\mathbf{S}_i)$. Suppose that $\{\mathbf{x}^{(n)},\; n = 0, 1, \cdots\}$ is a sequence of solutions generated by Algorithm 1. Then we have:

1) The following inequality holds for $n \geq 0$:

$$\hat{\varphi}^{(n)}(\mathbf{x}^{(n)}, \mathbf{S}^{(n)}) - \hat{\varphi}^{(n)}(\mathbf{x}^{(n+1)}, \mathbf{S}^{(n+1)}) \geq \frac{\rho_f + \rho_g}{2}\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2, \tag{37}$$

where ρ_f and ρ_g denote the convexity parameters of f_0 and g_0, respectively, i.e., $\rho_f, \rho_g > 0$ if f_0 and g_0 are strongly convex function and $\rho_f, \rho_g = 0$ otherwise.

2) If either f_0 or g_0 is strongly convex, i.e., $\rho_f + \rho_g > 0$, then $\Delta\mathbf{x}^{(n)}$ is a sufficient descent direction of (20) for all $n \geq 0$.

Proof: Please see Appendix A. ■

Subsequently, we assume that

Assumption 3 Either $f_0(\cdot)$ or $g_0(\cdot)$ is strongly convex.

The above assumption is needed to ensure $\Delta\mathbf{x}^{(n)}$ is a sufficient descent direction of (20) for all $n \geq 0$. To justify this assumption, let us consider a DC function $f(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x})$, then it is trivial to observe that $f(\mathbf{x}) = (f_1(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x}\|^2) - (f_2(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x}\|^2)$ for any given $\rho > 0$. Therefore, without loss of generality, we can always find a DC decomposition f_1, f_2 where both f_1 and f_2 are strongly convex.

[3] A similar example in convex optimization theory is that the Slater condition guarantees that the sufficient KKT conditions hold at some points for a convex problem.

The following theorem states the convergence properties of Algorithm 1:

Theorem 1: Let $\left\{\mathbf{x}^{(n)}\right\}$ be the solution sequence generated by Algorithm 1. Suppose (20) is feasible and A.1)–A.3) hold for (20), then one of the following scenarios applies:

1) Algorithm 1 terminates after a finite number of $\breve{n}$ iterations and $\mathbf{x}^{(\breve{n})}$ is a stationary point of (20);

2) Algorithm 1 generates an infinite sequence $\left\{\mathbf{x}^{(n)}\right\}$, then every limit point of $\{\mathbf{x}^{(n)}\}$ is a stationary point of (20).

Proof: Please see Appendix B. ■

Based on the above theorem, we can further obtain that the sequence of the objective function $\left\{\varphi(\mathbf{x}^{(n)})\right\}$ of (20) obtained by Algorithm 1 is also convergent.

IV. APPLICATION: SECURE MIMO AF RELAYING OPTIMIZATION

In this section, we apply the proposed penalized DC algorithm in the previous section to our secure MIMO AF relaying optimization problem (11d). We first show that the latter can be reformulated as a PSD DC program (20) by exploiting the so-called $\mathcal{S}$-procedure and by performing changes of variables. Subsequently, the penalized DC algorithm is adapted to solve the transformed optimization problem.

A. TRANSFORMATION OF (11d) INTO A PSD DC PROGRAM

The robust secure relaying optimization (11d) can be equivalently written as the following after substituting (4) and (7) into (11d),

$$\max_{\sigma_{\mathrm{S}},\mathbf{W},\boldsymbol{\Psi}} \; \frac{\sigma_{\mathrm{S}}^2|\mathbf{h}_2^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{h}_2^H\mathbf{W}\|^2+\mathbf{h}_2^H\boldsymbol{\Psi}\mathbf{h}_2+\sigma_{\mathrm{D}}^2} \tag{38a}$$

$$\text{s.t.} \quad \max_{\Delta g_{1k}\in\mathcal{G}_{1k}} \frac{\sigma_{\mathrm{S}}^2|g_{1k}|^2}{\sigma_{\mathrm{E},k}^2} \le \gamma,\; k\in\mathcal{K} \tag{38b}$$

$$\max_{\Delta\mathbf{g}_{2k}\in\mathcal{G}_{2k}} \frac{\sigma_{\mathrm{S}}^2|\mathbf{g}_{2k}^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{g}_{2k}^H\mathbf{W}\|^2+\mathbf{g}_{2k}^H\boldsymbol{\Psi}\mathbf{g}_{2k}+\sigma_{\mathrm{E},k}^2} \le \gamma, \quad k\in\mathcal{K} \tag{38c}$$

$$\sigma_{\mathrm{S}}^2\|\mathbf{W}\mathbf{h}_1\|^2+\sigma_{\mathrm{R}}^2\|\mathbf{W}\|_F^2+\mathrm{Tr}(\boldsymbol{\Psi}) \le P_{\mathrm{R}} \tag{38d}$$

$$\sigma_{\mathrm{S}}^2 \le P_{\mathrm{S}},\; \boldsymbol{\Psi}\succeq\mathbf{0}, \tag{38e}$$

where $\gamma = 2^{2\kappa R_d}-1$. Constraint (38b) can be equivalently rewritten as the following by exploiting the Cauchy-Schwarz inequality:

$$\sigma_{\mathrm{S}} \le \min_{k\in\mathcal{K}} \left\{ \frac{\gamma\sigma_{\mathrm{E},k}^2}{\left|\,|\hat{g}_{1k}|+\sqrt{\varepsilon_{1k}}\right|^2} \right\}. \tag{39}$$

Then to tackle the infiniteness associated with (38c), after some manipulations, we can rewrite (38c) as

$$\Delta\mathbf{g}_{2k}^H\boldsymbol{\Theta}(\mathbf{W},\boldsymbol{\Psi})\Delta\mathbf{g}_{2k}+2\,\mathrm{Re}\left\{\hat{\mathbf{g}}_{2k}^H\boldsymbol{\Theta}(\mathbf{W},\boldsymbol{\Psi})\Delta\mathbf{g}_{2k}\right\} + \hat{\mathbf{g}}_{2k}^H\boldsymbol{\Theta}(\mathbf{W},\boldsymbol{\Psi})\hat{\mathbf{g}}_{2k}-\gamma\sigma_{\mathrm{E},k}^2 \le 0,\; \forall\Delta\mathbf{g}_{2k}\in\mathcal{G}_{2k} \tag{40}$$

where we have defined $\boldsymbol{\Theta}(\mathbf{W},\boldsymbol{\Psi}) \triangleq \mathbf{W}\left(\sigma_{\mathrm{S}}^2\mathbf{h}_1\mathbf{h}_1^H-\gamma\sigma_{\mathrm{R}}^2\mathbf{I}_{N_{\mathrm{R}}}\right)\mathbf{W}^H-\gamma\boldsymbol{\Psi}$. As a popular technique of tackling the infiniteness in the robust optimization theory, we invoke the so-called $\mathcal{S}$-Procedure [38] for equivalently recasting (40) as

$$\mathbf{P}_k^H\boldsymbol{\Theta}(\mathbf{W},\boldsymbol{\Psi})\mathbf{P}_k-\boldsymbol{\Lambda}_k(\rho_k)\preceq\mathbf{0}, \tag{41}$$

where we have $\boldsymbol{\Lambda}_k(\rho_k)=\mathtt{blkdiag}(\rho_k\mathbf{I}_{N_{\mathrm{R}}},\gamma\sigma_{\mathrm{E},k}^2-\varepsilon_{2k}\rho_k)$ and $\mathbf{P}_k=[\mathbf{I}_{N_{\mathrm{R}}},\hat{\mathbf{g}}_{2k}]$ with $\mathtt{blkdiag}(\cdot,\cdot)$ denoting the construction of a block diagonal matrix from the input arguments.

To further transform (38) into a PSD DC program in the form of (20), let us introduce an auxiliary variable t. Plugging (40) and (41) back into (38), we obtain

$$\max_{\sigma_{\mathrm{S}},\mathbf{W},\boldsymbol{\Psi}} \; \frac{\sigma_{\mathrm{S}}^2|\mathbf{h}_2^H\mathbf{W}\mathbf{h}_1|^2}{t} \tag{42a}$$

$$\text{s.t.}\quad \sigma_{\mathrm{R}}^2\|\mathbf{h}_2^H\mathbf{W}\|^2+\mathbf{h}_2^H\boldsymbol{\Psi}\mathbf{h}_2+\sigma_{\mathrm{D}}^2\le t \tag{42b}$$

$$\sigma_{\mathrm{S}} \le \min_{k\in\mathcal{K}} \left\{ \frac{\gamma\sigma_{\mathrm{E},k}^2}{\left|\,|\hat{g}_{1k}|+\sqrt{\varepsilon_{1k}}\right|^2} \right\} \tag{42c}$$

$$\mathbf{P}_k^H\boldsymbol{\Theta}(\mathbf{W},\boldsymbol{\Psi})\mathbf{P}_k-\boldsymbol{\Lambda}_k(\rho_k)\preceq\mathbf{0},\; k\in\mathcal{K} \tag{42d}$$

$$\sigma_{\mathrm{S}}^2\|\mathbf{W}\mathbf{h}_1\|^2+\sigma_{\mathrm{R}}^2\|\mathbf{W}\|_F^2+\mathrm{Tr}(\boldsymbol{\Psi}) \le P_{\mathrm{R}} \tag{42e}$$

$$\sigma_{\mathrm{S}} \le \sqrt{P_{\mathrm{S}}},\; \boldsymbol{\Psi}\succeq\mathbf{0}. \tag{42f}$$

Observe that in the above formulation, the source power σ_{S} and relay AF matrix $\mathbf{W}$ are nonlinearly coupled in the objective (42a) and constraints (42d) and (42e). To transform (42) into a more convenient form, we introduce a new optimization variable $\mathbf{U}$, which is related to σ_{S} and $\mathbf{W}$ via the following bilinear matrix equality:

$$\mathbf{U}=\sigma_{\mathrm{S}}\mathbf{W}. \tag{43}$$

With the aid of (43), (42) can then be expressed as a PSD DC program with bilinear matrix equality constraint as defined in (16), i.e.,

$$\max_{\mathbf{x}} \; \frac{|\mathbf{h}_2^H\mathbf{U}\mathbf{h}_1|^2}{t} \tag{44a}$$

$$\text{s.t.}\quad \mathbf{P}_k^H\boldsymbol{\Theta}(\mathbf{W},\boldsymbol{\Psi})\mathbf{P}_k-\boldsymbol{\Lambda}_k(\rho_k)\preceq\mathbf{0},\; k\in\mathcal{K} \tag{44b}$$

$$\mathbf{U}=\sigma_{\mathrm{S}}\mathbf{W} \tag{44c}$$

$$\mathbf{x}\in\Omega, \tag{44d}$$

where $\mathbf{x}$ collectively denotes all the optimization variables (including both the original and auxiliary variables), i.e.,

$$\mathbf{x}\triangleq[\sigma_{\mathrm{S}},\mathrm{vec}(\mathbf{U})^T,\mathrm{vec}(\mathbf{W})^T,\mathrm{vec}(\boldsymbol{\Psi})^T,\boldsymbol{\rho}^T,t]^T, \tag{45}$$

and Ω is a compact convex subset defined as

$$\Omega\triangleq\{\mathbf{x}:(42\mathrm{b}),(42\mathrm{c}),(42\mathrm{e}),(42\mathrm{f})\}, \tag{46}$$

which can easily be represented as a finite number of LMIs by exploiting the techniques introduced in [39].

To tackle the bilinear matrix equality constraint (44c), we follow the procedure proposed in the previous section, i.e., exploit the results in Lemma 1, and conveniently convert (44c) into

$$\begin{bmatrix} \mathbf{Y}_1 & \mathbf{U} & \sigma_{\mathrm{S}}\mathbf{I}_{N_{\mathrm{R}}} \\ \mathbf{U}^H & \mathbf{Y}_2 & \mathbf{W}^H \\ \sigma_{\mathrm{S}}\mathbf{I}_{N_{\mathrm{R}}} & \mathbf{W} & \mathbf{I}_{N_{\mathrm{R}}} \end{bmatrix}\succeq\mathbf{0} \tag{47}$$

$$\mathrm{Tr}(\mathbf{Y}_1)-\mathrm{Tr}(\sigma_{\mathrm{S}}^2\mathbf{I}_{N_{\mathrm{R}}})\le 0, \tag{48}$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are auxiliary matrix variables with appropriate dimensions, (47) is an LMI, and (48) is a DC constraint (special case of PSD DC constraint with dimension one). Therefore, we can now embed the LMI (47) into Ω whilst preserving its convex structure. Additionally, the collection of optimization variables represented by $\mathbf{x}$ is augmented with the new auxiliary variables $\mathbf{Y}_1$ and $\mathbf{Y}_2$.

Finally, note that the matrix inequality constraint (44b) can expressed as a PSD DC constraint as follows:

$$\underbrace{\mathbf{P}_k^H\mathbf{U}\mathbf{h}_1\mathbf{h}_1^H\mathbf{U}^H\mathbf{P}_k - \boldsymbol{\Lambda}_k(\rho_k) - \gamma\mathbf{P}_k^H\boldsymbol{\Psi}\mathbf{P}_k}_{\boldsymbol{\mathcal{F}}_k(\cdot)} \\ -\underbrace{\gamma\sigma_{\mathrm{R}}^2\mathbf{P}_k^H\mathbf{W}\mathbf{W}^H\mathbf{P}_k}_{\boldsymbol{\mathcal{G}}_k(\cdot)} \preceq \mathbf{0},\ k\in\mathcal{K}, \tag{49}$$

where both $\boldsymbol{\mathcal{F}}_k(\cdot)$ and $\boldsymbol{\mathcal{G}}_k(\cdot)$ are PSD-convex mappings, which can be easily verified by Definition 1.

Based on the above derivations, we arrive at the following PSD DC program as defined (20):

$$\min_{\mathbf{x}} \ -\frac{|\mathbf{h}_2^H\mathbf{U}\mathbf{h}_1|^2}{t} \tag{50a}$$

$$\text{s.t. } \mathbf{P}_k^H\mathbf{U}\mathbf{h}_1\mathbf{h}_1^H\mathbf{U}^H\mathbf{P}_k - \boldsymbol{\Lambda}_k(\rho_k) - \gamma\mathbf{P}_k^H\boldsymbol{\Psi}\mathbf{P}_k \\ -\gamma\sigma_{\mathrm{R}}^2\mathbf{P}_k^H\mathbf{W}\mathbf{W}^H\mathbf{P}_k \preceq \mathbf{0},\ k\in\mathcal{K} \tag{50b}$$

$$\mathrm{Tr}(\mathbf{Y}_1) - \mathrm{Tr}(\sigma_{\mathrm{S}}^2\mathbf{I}_{N_{\mathrm{R}}}) \le 0 \tag{50c}$$

$$\mathbf{x}\in\Omega. \tag{50d}$$

B. PENALIZED PSD DC ALGORITHM FOR SECURE RELAYING DESIGN

For simplicity, let us denote

$$g_0(\mathbf{U},t) = -\frac{|\mathbf{h}_2^H\mathbf{U}\mathbf{h}_1|^2}{t} \tag{51}$$

$$g_1(\sigma_{\mathrm{S}}) = \mathrm{Tr}(\sigma_{\mathrm{S}}^2\mathbf{I}_{N_{\mathrm{R}}}) = N_{\mathrm{R}}\sigma_{\mathrm{S}}^2. \tag{52}$$

The algorithm designed for (50) is described as Algorithm 2, where $\hat{g}_0(\cdot;\cdot)$, $\hat{g}_1(\cdot;\cdot)$ and $\hat{\boldsymbol{\mathcal{G}}}_k(\cdot;\cdot)$ denote the first-order approximations of their corresponding functions/mapping around a solution from the previous iteration:

$$\hat{g}_0(\mathbf{U},t;\mathbf{U}^{(n),t^{(n)}}) \\ = \frac{\left|\mathbf{h}_2^H\mathbf{U}^{(n)}\mathbf{h}_1\right|^2}{t^{(n)}} - \frac{\left|\mathbf{h}_2^H\mathbf{U}^{(n)}\mathbf{h}_1\right|^2}{(t^{(n)})^2}\left(t-t^{(n)}\right) \\ + \frac{1}{t^{(n)}}2\,\mathrm{Re}\left\{\mathbf{h}_2^H\mathbf{U}^{(n)}\mathbf{h}_1\mathbf{h}_1^H\left(\mathbf{U}-\mathbf{U}^{(n)}\right)^H\mathbf{h}_2\right\} \tag{53}$$

$$\hat{g}_1(\sigma_{\mathrm{S}};\sigma_{\mathrm{S}}^{(n)}) = N_{\mathrm{R}}(\sigma_{\mathrm{S}}^{(n)})^2 - 2N_{\mathrm{R}}\sigma_{\mathrm{S}}^{(n)}\left(\sigma_{\mathrm{S}}+\sigma_{\mathrm{S}}^{(n)}\right) \tag{54}$$

$$\hat{\boldsymbol{\mathcal{G}}}_k(\mathbf{W};\mathbf{W}^{(n)}) = \gamma\sigma_{\mathrm{R}}^2\mathbf{P}_k^H\mathbf{W}^{(n)}\left(\mathbf{W}^{(n)}\right)^H\mathbf{P}_k \\ +2\gamma N_{\mathrm{R}}\,\mathrm{Re}\left(\mathbf{P}_k^H\mathbf{W}^{(n)}\left(\mathbf{W}-\mathbf{W}^{(n)}\right)^H\mathbf{P}_k\right). \tag{55}$$

We now briefly analyze the theoretical complexity of solving each sub-problem in (56). Since (56) is a standard SDP, its complexity mainly depends on the number of optimization variables and the number of semidefinite cone constraints. It is not difficult to verify (56) involves on the order of $\mathcal{O}(N_{\mathrm{R}}^2 + N_{\mathrm{R}} + K + 1)$ optimization variables and K semidefinite cone constraints of dimension $(N_{\mathrm{R}}+1)^2$ Therefore, as analyzed in [39], (56) can be solved at a worst case complexity, which is on the order of $\mathcal{O}((N_{\mathrm{R}}^2 + N_{\mathrm{R}} + K + 1)^2(N_{\mathrm{R}}+1)^2$.

Algorithm 2 Penalized PSD DC Algorithm for Secure Relaying Design

```
Intialization: An initial point $\mathbf{x}^{(0)} \in \Omega$, $\tau^{(0)} > 0$, $\delta_1 > 0$
  and $\delta_2 > 0$. Set $n = 0$.
  repeat
    Compute $\mathbf{x}^{(n+1)}$ by solving the convex sub-problem:
      $\min_{\mathbf{x}} \ -\hat{g}_0(\mathbf{U},t;\mathbf{U}^{(n)},t^{(n)}) + \tau^{(n)}(s + \sum_{k=1}^K \mathrm{Tr}(\mathbf{S}_k))$   (56a)
      s.t. $\boldsymbol{\mathcal{F}}_k - \hat{\boldsymbol{\mathcal{G}}}_k(\mathbf{W},\mathbf{W}^{(n)}) \preceq \mathbf{S}_k,\ k\in\mathcal{K}$   (56b)
           $\mathrm{Tr}(\mathbf{Y}_1) - \hat{g}_1(\sigma_{\mathrm{S}},\sigma_{\mathrm{S}}^{(n)}) \le s$   (56c)
           $\mathbf{x}\in\Omega.$   (56d)
    Update $\tau$ via (30);
    Update iteration: $n \leftarrow n+1$
  until Termination criterion is satisfied or a maximum
  number of iterations are reached
```

Before leaving this section, the following remarks are of interests:

Remark 3 (On the initialization of Algorithm 2): Since Ω defined in (46) is a compact convex subset, we are able to efficiently exploit its bounded structure, and conveniently select a feasible initialization, e.g.,

$$\sigma_{\mathrm{S}}^{(0)} = \min\left\{\sqrt{P_{\mathrm{S}}}, \min_{k\in\mathcal{K}}\left\{\frac{\gamma\sigma_{\mathrm{E},k}^2}{\left||\hat{g}_{1k}| + \sqrt{\varepsilon_{1k}}\right|^2}\right\}\right\} - \epsilon$$

$$\mathbf{W}^{(0)} = \left(\frac{P_{\mathrm{R}}}{(\sigma_{\mathrm{S}}^{(0)})^2\|\mathbf{h}_1\|^2 + \sigma_{\mathrm{R}}^2N_{\mathrm{R}}}\right)^{\frac{1}{2}}\mathbf{I}_{N_{\mathrm{R}}}$$

$$\mathbf{U}^{(0)} = \sigma_{\mathrm{S}}^{(0)}\mathbf{W}^{(0)}.$$

where ϵ is a small positive number. □

Remark 4 (Extension to the case of MRC): We show that the proposed penalized DC algorithm is also applicable to the case where eves adopt a more complicated receive MRC scheme for decoding the messages from the legitimate UEs. With MRC, the mutual information leakage to the eves can now be given by

$$\mathtt{C}_{\mathrm{E},k}(\sigma_{\mathrm{S}},\mathbf{W},\boldsymbol{\Psi}) \\ = \frac{1}{2}\log_2\left(1+\frac{\sigma_{\mathrm{S}}^2|\mathbf{g}_{1k}|^2}{\sigma_{\mathrm{E},k}^2} \\ +\frac{\sigma_{\mathrm{S}}^2|\mathbf{g}_{2k}^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{g}_{2k}^H\mathbf{W}\|^2 + \mathbf{g}_{2k}^H\boldsymbol{\Psi}\mathbf{g}_{2k} + \sigma_{\mathrm{E},k}^2}\right). \tag{57}$$

We adopt a rate-splitting approach, i.e., we introduce a pair of weights (γ_1,γ_2) with $\gamma_1+\gamma_2=\gamma$ as defined below (38), and the robust secrecy constraint can be subsequently formulated as

$$\begin{cases} \max\limits_{\Delta g_{1k}\in\mathcal{G}_{1k}} \frac{\sigma_{\mathrm{S}}^2|g_{1k}|^2}{\sigma_{\mathrm{E},k}^2} \leq \gamma_1,\ k\in\mathcal{K} \\ \max\limits_{\Delta \mathbf{g}_{2k}\in\mathcal{G}_{2k}} \frac{\sigma_{\mathrm{S}}^2|\mathbf{g}_{2k}^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{g}_{2k}^H\mathbf{W}\|^2+\mathbf{g}_{2k}^H\mathbf{\Psi}\mathbf{g}_{2k}+\sigma_{\mathrm{E},k}^2} \leq \gamma_2,\ k\in\mathcal{K} \end{cases}. \tag{58}$$

It can be observed that in the case of MRC, the robust secrecy constraints have a similar form to that of the SC by prefixing a pair of weights $(\beta, 2R_d - \beta)$. Therefore, given different values of β, we can obtain a set of solutions $(\sigma_{\mathrm{S}}, \mathbf{W}, \mathbf{\Psi})$ using Algorithm 2. Within such set of solution, the best solution can be achieved by the one that attains the maximum $\mathtt{SINR}_{\mathrm{D}}$ in the objective function. □

Remark 5 (Extension to secrecy-capacity based optimization): An alternative approach to improve secrecy from information theoretical perspective is to maximize the *worst-case* achievable secrecy rate of the relaying network subject to power constraints, which can mathematically expressed as

$$\max_{\sigma_{\mathrm{S}},\mathbf{W},\mathbf{\Psi}}\ R_S \quad \text{s.t. (38d) and (38e)}, \tag{59}$$

where R_S denotes the worst-case achievable secrecy rate, which is given by

$$\begin{aligned} R_S = \Bigg\{ &\log_2\left(1+\frac{\sigma_{\mathrm{S}}^2|\mathbf{h}_2^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{h}_2^H\mathbf{W}\|^2+\mathbf{h}_2^H\mathbf{\Psi}\mathbf{h}_2+\sigma_{\mathrm{D}}^2}\right) \\ &-\max_{k\in\mathcal{K}}\ \max_{\substack{\Delta g_{1k}\in\mathcal{G}_{1k}\\ \Delta\mathbf{g}_{2k}\in\mathcal{G}_{2k}}} \Bigg[\log_2\left(1+\frac{\sigma_{\mathrm{S}}^2|g_{1k}|^2}{\sigma_{\mathrm{E},k}^2}\right) \\ &+\log_2\left(1+\frac{\sigma_{\mathrm{S}}^2|\mathbf{g}_{2k}^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{g}_{2k}^H\mathbf{W}\|^2+\mathbf{g}_{2k}^H\mathbf{\Psi}\mathbf{g}_{2k}+\sigma_{\mathrm{E},k}^2}\right)\Bigg]\Bigg\}. \end{aligned} \tag{60}$$

By introducing a few auxiliary variables $\mathbf{t} = [t_1, t_2, t_3]^T \geq 0$, we can re-express (59) as

$$\min_{\sigma_{\mathrm{S}},\mathbf{W},\mathbf{\Psi},\mathbf{t}}\ \log_2(1+t_1) - \big(\log_2(1+t_2)+\log_2(1+t_3)\big) \tag{61a}$$

$$\text{s.t.}\ \frac{\sigma_{\mathrm{S}}^2|\mathbf{h}_2^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{h}_2^H\mathbf{W}\|^2+\mathbf{h}_2^H\mathbf{\Psi}\mathbf{h}_2+\sigma_{\mathrm{D}}^2} \leq t_1 \tag{61b}$$

$$\max_{\Delta g_{1k}\in\mathcal{G}_{1k}} \frac{\sigma_{\mathrm{S}}^2|g_{1k}|^2}{\sigma_{\mathrm{E},k}^2} \leq \gamma_1,\ k\in\mathcal{K} \tag{61c}$$

$$\max_{\Delta\mathbf{g}_{2k}\in\mathcal{G}_{2k}} \frac{\sigma_{\mathrm{S}}^2|\mathbf{g}_{2k}^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{g}_{2k}^H\mathbf{W}\|^2+\mathbf{g}_{2k}^H\mathbf{\Psi}\mathbf{g}_{2k}+\sigma_{\mathrm{E},k}^2} \leq \gamma_2 \tag{61d}$$

$$\text{(38d) and (38e)}. \tag{61e}$$

Since $-\log_2(\cdot)$ is a convex function, the objective (61a) is simply a DC function. Additionally, constraints (61b)–(61d) are in the forms similar to those of (38a)–(38c), respectively. Hence, the proposed penalized PSD DC algorithm in Algorithm 1 can be accordingly adapted to the secrecy rate maximization problem (61) following the transformation similar to Section IV–B.

V. BENCHMARKER: SDR-BASED EXHAUSTIVE SEARCH METHOD

To benchmark the proposed penalized DC algorithm, in this section we derive an SDR-based approach that yields an upper-bound for the robust secrecy problem (11d), however, at the expense of higher computational complexity.

It is in general challenging to jointly optimize the tuple of $(\sigma_{\mathrm{S}}, \mathbf{W}, \mathbf{\Psi})$ due to its non-convex nature and therefore, we can consider a sub-problem of (11d) solving for the optimal pair $(\mathbf{W}, \mathbf{\Psi})$, while temporarily fixing the value of σ_{S}. Substituting the expression of $\mathtt{C}_{\mathrm{E},k}$ in (7) into (11b) and neglecting the terms independent of $(\mathbf{W}, \mathbf{\Psi})$, we arrive at the sub-problem (62), shown at the top of the next page, where $\tau(\sigma_{\mathrm{S}})$ denotes its objective value, which depends on the value of σ_{S}. With the aid of (62), the original problem (11d) can equivalently be expressed as

$$\max_{\sigma_{\mathrm{S}}}\ \tau(\sigma_{\mathrm{S}}) \quad \text{s.t.}\ \ 0 \leq \sigma_{\mathrm{S}} \leq \bar{\sigma}_{\mathrm{S}}, \tag{63}$$

where σ_{S} is lower bounded by zero, while its upper bound $\bar{\sigma}_{\mathrm{S}}$ is given by [c.f. (39)]

$$\bar{\sigma}_{\mathrm{S}} = \min\left\{\sqrt{P_{\mathrm{S}}}, \min_{k\in\mathcal{K}}\left\{\frac{\gamma\sigma_{\mathrm{E},k}^2}{\big||\hat{g}_{1k}|+\sqrt{\varepsilon_{1k}}\big|^2}\right\}\right\}, \tag{64}$$

The reformulated problem in (63) leads to a simpler single-variable optimization problem defined over the interval $[0,\ \bar{\sigma}_{\mathrm{S}}]$. Assuming that $\tau(\sigma_{\mathrm{S}})$ can be evaluated at any feasible σ_{S}, a one-dimensional exhaustive search procedure can be invoked for finding the global optimum of (11d). Let us now focus our attention on computing $\tau(\sigma_{\mathrm{S}})$ for a given feasible σ_{S}, which however requires solving the non-convex sub-problem (62). The solution to (62) will be addressed in the following.

Recall that the infiniteness of the constraint in (62b) can be tackled by the $\mathcal{S}$-procedure [c.f. (40), (41)], which leads to the following equivalent reformulation:

$$\mathbf{\Lambda}_k(\rho_k) - \mathbf{P}_k^H\mathbf{\Theta}(\mathbf{W},\mathbf{\Psi})\mathbf{P}_k \succeq \mathbf{0}. \tag{65}$$

Replacing (62b) by (65), the sub-problem in $(\mathbf{W}, \mathbf{\Psi})$ of (62) can now be rewritten in a *finite* form:

$$\max_{\mathbf{W},\mathbf{\Psi}}\ \frac{\sigma_{\mathrm{S}}^2|\mathbf{h}_2^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{h}_2^H\mathbf{W}\|^2+\mathbf{h}_2^H\mathbf{\Psi}\mathbf{h}_2+\sigma_{\mathrm{D}}^2} \tag{66a}$$

$$\text{s.t.}\ \sigma_{\mathrm{S}}^2\|\mathbf{W}\mathbf{h}_1\|^2+\sigma_{\mathrm{R}}^2\|\mathbf{W}\|_F^2+\mathrm{Tr}(\mathbf{\Psi}) \leq P_{\mathrm{R}} \tag{66b}$$

$$\mathbf{\Lambda}_k(\rho_k) - \mathbf{P}_k^H\mathbf{\Theta}(\mathbf{W},\mathbf{\Psi})\mathbf{P}_k \succeq \mathbf{0},\ k\in\mathcal{K} \tag{66c}$$

$$\mathbf{\Psi} \succeq \mathbf{0}. \tag{66d}$$

The above transformed formulation is still non-convex and to proceed, we have to transform it into an appropriate formulation, where the SDR is applicable. Let us define $\mathbf{w} = \mathrm{vec}(\mathbf{W})$ and $\mathbf{X} = \mathbf{w}\mathbf{w}^H$. Interestingly, after some tedious matrix manipulations, which have been relegated to Appendix XII, we are now able to rewrite (66) in a form, which only involves the linear terms of $\mathbf{X}$ and $\mathbf{\Psi}$. The results are summarized in the following proposition:

$$\tau(\sigma_{\mathrm{S}}) \triangleq \max_{\mathbf{W},\boldsymbol{\Psi}\succeq\mathbf{0}} \frac{\sigma_{\mathrm{S}}^2|\mathbf{h}_2^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{h}_2^H\mathbf{W}\|^2+\mathbf{h}_2^H\boldsymbol{\Psi}\mathbf{h}_2+\sigma_{\mathrm{D}}^2} \tag{62a}$$

$$\text{s.t. (11c) and } \log_2\left(1+\frac{\sigma_{\mathrm{S}}^2|\mathbf{g}_{2k}^H\mathbf{W}\mathbf{h}_1|^2}{\sigma_{\mathrm{R}}^2\|\mathbf{g}_{2k}^H\mathbf{W}\|^2+\mathbf{g}_{2k}^H\boldsymbol{\Psi}\mathbf{g}_{2k}+\sigma_{\mathrm{E},k}^2}\right)\le\kappa R_d,\ \forall\Delta\mathbf{g}_{2k}\in\mathcal{G}_{2k}, k\in\mathcal{K}. \tag{62b}$$

Proposition 2 Define

$$\mathbf{Q}_0=\sigma_{\mathrm{S}}^2(\mathbf{h}_1^*\mathbf{h}_1^T)\otimes(\mathbf{h}_2\mathbf{h}_2^H) \tag{67}$$

$$\mathbf{Q}_1=\sigma_{\mathrm{R}}^2\mathbf{I}_{N_{\mathrm{R}}}\otimes(\mathbf{h}_2\mathbf{h}_2^H) \tag{68}$$

$$\mathbf{Q}_2=\sigma_{\mathrm{S}}^2(\mathbf{h}_1^*\mathbf{h}_1^T)\otimes\mathbf{I}_{N_{\mathrm{R}}}+\sigma_{\mathrm{R}}^2\mathbf{I}_{N_{\mathrm{R}}^2} \tag{69}$$

$$\mathbf{Q}_3(\mathbf{X},\boldsymbol{\Psi})=\sigma_{\mathrm{S}}^2\mathbf{H}_1\mathbf{X}\mathbf{H}_1^H-\gamma\sigma_{\mathrm{R}}^2\sum_{l=1}^{N_{\mathrm{R}}}\mathbf{E}_l\mathbf{X}\mathbf{E}_l^H-\gamma\boldsymbol{\Psi}, \tag{70}$$

where $\mathbf{Q}_3(\cdot)$ is a linear mapping of $\mathbf{X}$ and $\boldsymbol{\Psi}$ with $\mathbf{H}_1=\mathbf{h}_1^T\otimes\mathbf{I}_{N_{\mathrm{R}}}$ and $\mathbf{E}_l=\left[\mathbf{0}_{N_{\mathrm{R}}\times(l-1)N_{\mathrm{R}}},\mathbf{I}_{N_{\mathrm{R}}},\mathbf{0}_{N_{\mathrm{R}}\times(N_{\mathrm{R}}-l)N_{\mathrm{R}}}\right]$. Then problem (66) can equivalently be rewritten in the following form:

$$\max_{\mathbf{X},\boldsymbol{\Psi},\boldsymbol{\rho}} \frac{\mathrm{Tr}(\mathbf{Q}_0\mathbf{X})}{\mathrm{Tr}(\mathbf{Q}_1\mathbf{X})+\mathrm{Tr}(\mathbf{h}_2\mathbf{h}_2^H\boldsymbol{\Psi})+\sigma_{\mathrm{D}}^2} \tag{71a}$$

$$\text{s.t. } \mathrm{Tr}(\mathbf{Q}_2\mathbf{X})+\mathrm{Tr}(\boldsymbol{\Psi})\le P_{\mathrm{R}} \tag{71b}$$

$$\boldsymbol{\Lambda}_k(\rho_k)-\mathbf{P}_k^H\mathbf{Q}_3(\mathbf{X},\boldsymbol{\Psi})\mathbf{P}_k\succeq\mathbf{0},\ k\in\mathcal{K} \tag{71c}$$

$$\mathbf{X}\succeq\mathbf{0},\ \boldsymbol{\Psi}\succeq\mathbf{0},\ \mathrm{Rank}(\mathbf{X})=1. \tag{71d}$$

Upon neglecting the non-convex rank-one constraint in (71d), (71) is relaxed to a so-called fractional SDP, which can further be transformed into a standard SDP via the Charnes-Cooper transformation [40]. Specifically, by introducing an auxiliary variable $s>0$, and defining $\overline{\mathbf{X}}=s\mathbf{X}$, $\overline{\boldsymbol{\Psi}}=s\boldsymbol{\Psi}$ and $\overline{\boldsymbol{\rho}}=s\boldsymbol{\rho}$, (71) is conveniently recast as

$$\max_{\overline{\mathbf{X}},\overline{\boldsymbol{\Psi}},\overline{\boldsymbol{\rho}},s>0} \mathrm{Tr}(\mathbf{Q}_0\overline{\mathbf{X}}) \tag{72a}$$

$$\text{s.t. } \mathrm{Tr}(\mathbf{Q}_1\overline{\mathbf{X}})+\mathrm{Tr}(\mathbf{h}_2\mathbf{h}_2^H\overline{\boldsymbol{\Psi}})+s\sigma_{\mathrm{D}}^2\le 1 \tag{72b}$$

$$\mathrm{Tr}(\mathbf{Q}_2\overline{\mathbf{X}})+\mathrm{Tr}(\overline{\boldsymbol{\Psi}})\le sP_{\mathrm{R}} \tag{72c}$$

$$\boldsymbol{\Lambda}_k(\overline{\rho}_k)-\mathbf{P}_k^H\mathbf{Q}_3(\overline{\mathbf{X}},\overline{\boldsymbol{\Psi}})\mathbf{P}_k\succeq\mathbf{0},\ k\in\mathcal{K} \tag{72d}$$

$$\overline{\mathbf{X}}\succeq\mathbf{0},\ \overline{\boldsymbol{\Psi}}\succeq\mathbf{0}. \tag{72e}$$

Interestingly, (72) now becomes a convex SDP, which is efficiently solvable by generic optimization tools such as `SeDuMi` [41] and `MOSEK` [42] relying on interior-point methods [43]. We remark that (72) and the rank-relaxed version of (71) are equivalent in the sense that the optimal solution $\mathbf{X}^*$ to (71) after rank-one relaxation can be retrieved by the optimal solution $(\overline{\mathbf{X}}^*, s^*)$ to (72), i.e., $\mathbf{X}^*=\frac{\overline{\mathbf{X}}^*}{s^*}$, and the resultant objective values of the two problems are equivalent.

After obtaining the rank-relaxed solution $\mathbf{X}^*$, a natural question arises as to how good a solution is $\mathbf{X}^*$, i.e., does it satisfy the rank-one optimality condition of (71)? Answering these questions directly from the formulation of (72) is still an open problem in the literature. To overcome this difficulty, we follow an approach similar to [21]. Specifically, denoting the objective value of (72) by $\tau_{\mathrm{relax}}^*(\sigma_{\mathrm{S}})$, we consider the following power minimization problem:

$$\min_{\mathbf{X},\boldsymbol{\Psi},\boldsymbol{\rho}} \mathrm{Tr}(\mathbf{Q}_2\mathbf{X}) \tag{73a}$$

$$\text{s.t. } \frac{\mathrm{Tr}(\mathbf{Q}_0\mathbf{X})}{\mathrm{Tr}(\mathbf{Q}_1\mathbf{X})+\mathrm{Tr}(\mathbf{h}_2\mathbf{h}_2^H\boldsymbol{\Psi})+\sigma_{\mathrm{D}}^2}\ge\tau_{\mathrm{relax}}^*(\sigma_{\mathrm{S}}) \tag{73b}$$

$$\mathrm{Tr}(\mathbf{Q}_2\mathbf{X})+\mathrm{Tr}(\boldsymbol{\Psi})\le P_{\mathrm{R}} \tag{73c}$$

$$\boldsymbol{\Lambda}_k(\rho_k)-\mathbf{P}_k^H\mathbf{Q}_3(\mathbf{X},\boldsymbol{\Psi})\mathbf{P}_k\succeq\mathbf{0},\ k\in\mathcal{K} \tag{73d}$$

$$\mathbf{X}\succeq\mathbf{0},\ \boldsymbol{\Psi}\succeq\mathbf{0}. \tag{73e}$$

Observe that (73) is also a standard SDP and therefore it is readily solvable by existing optimization tools. Furthermore, its specific structure allows us to obtain the following useful results, based on which we are able to retrieve an optimal rank-one solution of (71).

Proposition 3 Let us denote the optimal solution of (73) by $(\mathbf{X}^o,\boldsymbol{\Psi}^o,\boldsymbol{\rho}^o)$. Assuming suitable constraint qualification of (73), $(\mathbf{X}^o,\boldsymbol{\Psi}^o,\boldsymbol{\rho}^o)$ is also an optimal solution of (71), i.e., $\mathbf{X}^o$ must be of rank one.

Proof: Please see appendix F. ■

In summary, obtaining an optimal solution of (63) now consists of two steps: *1)* solve the rank-relaxed SDP (72) and obtain the largest $\tau_{\mathrm{relax}}(\sigma_{\mathrm{S}})$ by exhaustive search over σ_{S}; *2)* solve the power minimization problem (73) based on $\tau_{\mathrm{relax}}(\sigma_{\mathrm{S}})$. Since the rank-one optimality condition of $\mathbf{X}^o$ is guaranteed, the optimal AF matrix $\mathbf{W}^o$ can be retrieved by the rank-one decomposition of $\mathbf{X}^o$, i.e., $\mathbf{X}^o=\mathbf{x}^o(\mathbf{x}^o)^H$ and subsequently converting $\mathbf{x}^o$ to $\mathbf{W}^o$ via the vector-matrix reshaping.

We should point out that solving (63) requires performing an exhaustive search for σ_{S} over $[0,\ \bar{\sigma}_{\mathrm{S}}]$. In each step, we have to solve the SDP (72), which involves on the order of $\mathcal{O}(N_{\mathrm{R}}^4+N_{\mathrm{R}}^2+1)$ optimization variables and K semidefinite cone constraints of dimension $(N_{\mathrm{R}}+1)^2$. Therefore, it can be solved at a *worst-case* complexity, which is on the order of $\mathcal{O}\left(K(N_{\mathrm{R}}^4+N_{\mathrm{R}}^2+1)^2(N_{\mathrm{R}}+1)^2\right)$ [39]. As compared to the complexity of the proposed penalized DC algorithm (see, e.g., analysis below Algorithm 2), The associated computational cost escalates significantly faster as the size of the relay antenna array and the number of `eves` increase, which may become computationally prohibitive in practical problems.

VI. NUMERICAL EXAMPLES

The efficacy of the proposed solutions to the robust secure relaying problem is verified by a few numerical examples. In all simulations, all the coefficients of the legitimate channels $\mathbf{h}_1$ and $\mathbf{h}_2$, and the estimated `eves`' channels $\{\hat{g}_{1,k}\}$ and $\{\hat{\mathbf{g}}_{2,k}\}$ are generated following identically and independently

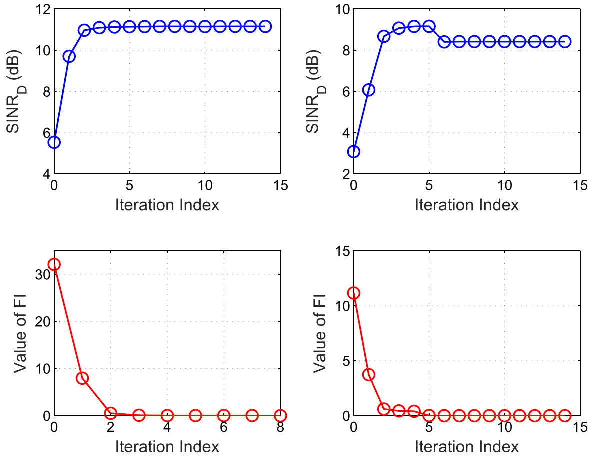


FIGURE 2. Convergence behavior of Algorithm 1. Left set of sub-figures: The first case. Right set of sub-figures: The second case.

distributed (i.i.d.) complex circular Gaussian distribution with zero-mean and unit-variance. Equal radii are assumed for all $\Delta g_{1,k}$ and for all $\Delta \mathbf{g}_{2,k}$, i.e., $\varepsilon_{1,k} = \varepsilon_1$ and $\varepsilon_{2,k} = \varepsilon_2$ for all k. The power budget of S is normalized to one and we set higher power budget for R with $P_{\mathrm{R}} = 2$. It is also assumed that an antenna array of size $N_{\mathrm{R}} = 3$ is employed by R. The noise variances are $\sigma_{\mathrm{R}}^2 = 0.05$, $\sigma_{\mathrm{D}}^2 = 0.05$ and $\sigma_{\mathrm{E},k}^2 = 0.01$ $\forall k$. The above parameters are fixed unless otherwise explicitly stated. In all figures, we denote the proposed penalized DC algorithm in Section IV by ''Proposed P-DCA'' and the derived benchmarker in Section V by ''SDR+Search''.

1) CONVERGENCE

We first study the convergence behavior of Algorithm 1. We simulate 200 channel realizations and among which, two classes of behaviors are observed. A representative case for each class is then plotted in the left and right parts of Fig. 2. In each case, the top sub-figure shows the convergence of the achieved SINR at D whilst the bottom sub-figure plots the evolution of the FI. The first case shows a behavior similar to conventional DC algorithm. The second example shows a more interesting behavior where the algorithm begins with an infeasible point and in the first few iterations, the algorithm targets finding a region (still infeasible) with larger objective function. As the penalty terms gradually play more important roles, more emphasis will be on finding a feasible point near the above located region. Therefore, the value of objective function drops since the feasibility has to be enforced now. Finally, the SINR remains approximately the same because a stationary point is achieved. The convergence behavior is consistent with the discussions and proof in Section IV.

2) SECRECY

To evaluate the secrecy of relaying transmission achieved by the proposed solutions, i.e., how consistently the robust secrecy constraints (11b) can be satisfied, we follow a probabilistic approach similar to [44, Sec. VI–B]. In this example, the coefficients of $\Delta\hat{g}_{1k}$ and $\Delta\hat{\mathbf{g}}_{2k}$ are generated by i.i.d. zero-mean complex circular Gaussian distribution with variance $\sigma_h^2 = 0.05$. The radii of uncertainty regions in (9) and (10) are then determined by $\varepsilon_1 = \sigma_h^2 \times \mathtt{gammaincinv}(\mathrm{Pr}, 0.5)$ and $\varepsilon_2 = \sigma_h^2 \times \mathtt{gammaincinv}(\mathrm{Pr}, 0.5N_{\mathrm{R}}^2)$ where $\mathtt{gammaincinv}(\cdot)$ is the inverse of incomplete gamma function defined in MATLAB and Pr is a predefined bounding probability, say, $\mathrm{Pr} = 95\%$, c.f. [44, (61)]. The empirical cumulative distribution functions (CDFs) of mutual information leakage at both eves are shown in Fig 3. Both the proposed solutions ensures that the mutual information leakage never exceeds the data rate of legitimate UEs whilst the non-robust design leads to a frequent violation of the secrecy constraints, namely for more than 20% of the realizations. Considering the practical MCS with finite coding block length, a proper selection of κ would lead to sufficiently high block error rate (BLER) at eves. Although the proposed method can prevent the eves from perfectly decoding the information signals, we need to point out the use of the secrecy constraints does not guarantee perfect secrecy from the information theoretical perspective. However, we can view our design as a means to cause additional confusion to eves.

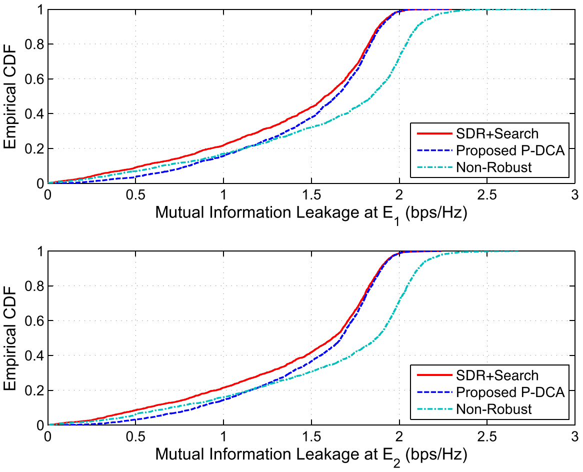


FIGURE 3. Empirical CDFs of mutual information leakage at eves. The legitimate S is transmitting at $R_d = 2$ bps/Hz.

3) RELIABILITY

Having verified the secrecy of the proposed solutions, we now compare the transmission reliability in terms of the achieved SINR at D. In Fig. 4, $\mathrm{SINR}_{\mathrm{D}}$ for a set of 50 independent experiments are plotted. The curve labeled ''Nullspace Beamforming'' refers to the method where R first nullifies eves' reception by first projecting its received signal onto the null space of $[\hat{\mathbf{g}}_{2,1}, \cdots, \hat{\mathbf{g}}_{2,K}]$ and then performs AF relaying. Therefore, the method is only applicable when $N_{\mathrm{R}} > K$. Two cases $K = 2$ and $K = 4$ are considered. In both cases, we observe that the performance of the proposed penalized DC algorithm is very close to the SDR-based benchmarker. In the case of $K = 2$, the proposed solution significantly outperforms the nullspace beamforming method.

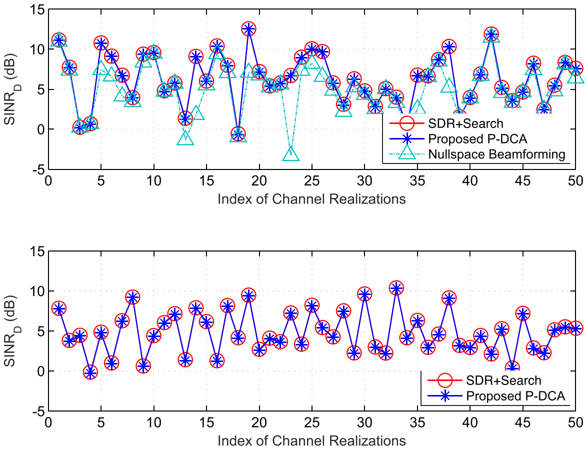


FIGURE 4. **Achieved SINR at** D. **Top sub-figure:** $K = 2$ eves. **Bottom sub-figure:** $K = 4$ eves.

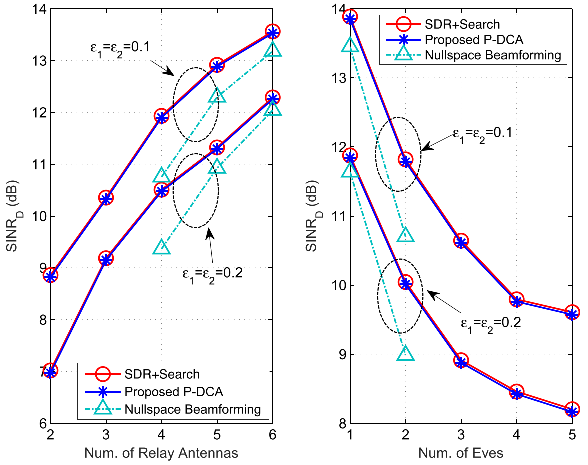


FIGURE 5. **Achieved SINR at** D. **Left sub-figure:** $\mathrm{SINR_D}$ **versus** N_R. $K = 3$ eves **are considered. Right sub-figure:** $\mathrm{SINR_D}$ **versus** K.

We then study how different system configurations impact the achieved SINR by different approaches. In the left sub-figure of Fig. 5, the achieved SINR of the proposed solutions and the nullspace beamforming is plotted as a function of the number of antenna elements employed at R. Two sizes of uncertainty regions are considered with $\varepsilon_1 = \varepsilon_2 = 0.1$ and $\varepsilon_1 = \varepsilon_2 = 0.2$. In both scenarios, the achieved SINR monotonically increases as N_R increases due to the higher diversity one can exploit from the antenna array. Again, both the proposed solutions consistently exhibit better performance than the nullspace beamforming. Notice also when more channel uncertainties are now present ($\varepsilon_1 = \varepsilon_2 = 0.2$), the legitimate UEs are confined to relatively low transmission power to satisfy the robust secrecy constraints, leading to lower received SINR at D. In the right sub-figure of Fig. 5, the impact of different number of eves on the achieved SINR is assessed. The SINR monotonically decreases when there are more eves around and therefore, the legitimate UEs have to lower their transmission power to prevent the information leakage more carefully. For completeness, we also investigate how robustly the proposed solutions can behave against the

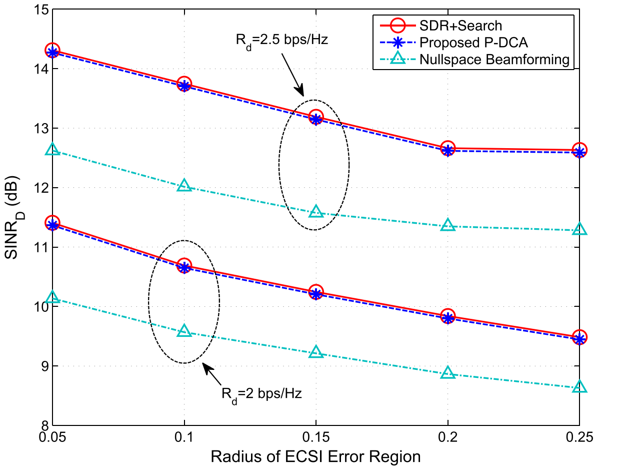


FIGURE 6. **Achieved SINR at** D **as a function of size of uncertainty region. Two data rates of legitimate UEs are considered, namely,** $R_d = 2$**bps/Hz and** $R_d = 2.5$**bps/Hz.**

TABLE 1. **Average solver time (in seconds) for different algorithms.**

			Num. of Relay Ant. N_R				
		Alg.	2	3	4	5	6
Num. of eves K	2	SDR	1.31	4.44	22.44	111.26	553.11
		P-DCA	0.69	1.42	2.58	4.73	8.38
	3	SDR	1.59	5.95	28.61	141.54	680.90
		P-DCA	0.88	1.80	3.45	6.52	11.31
	4	SDR	1.83	7.14	33.29	165.29	798.69
		P-DCA	1.05	2.24	4.60	8.01	14.09
	5	SDR	2.19	8.38	41.23	203.20	924.32
		P-DCA	1.20	2.66	5.43	9.95	17.12

ECSI errors by varying the sizes of the channel uncertainty regions. Again, the results are as expected and showing the superiority of our proposed solutions.

4) COMPUTATIONAL COMPLEXITY

Last but not least, we need to justify the lower complexity of the proposed penalized DC algorithm as compared to the SDR-based benchmarker proposed in Section V. The averaged solver time over 100 independent realizations is shown in Table 1 for different values of N_R and K. It is observed that the solver time for the SDR approach scales very fast with increases in N_R and K, which is consistent with the worst-case complexity analysis in Section V. In the meantime, the solver time of the proposed penalized DC algorithm increases more slowly compared to the former.

VII. CONCLUSIONS

Robust design of secure MIMO relaying in the presence of multiple eves was studied. We jointly optimized the power of S, the AF matrix and covariance of AN at R to maximize the received SINR at D while imposing a set of mutual information leakage-based secrecy constraints. With only imperfect ECSI, the resultant problem has been shown to be non-convex and challenging. A computationally efficient sub-optimal solution relying on the new penalized DC algorithmic framework was developed. This algorithm is capable of finding a stationary solution to a general non-convex SDP representable by a PSD DC program. The latter can be efficiently solved by the penalized DC algorithm without finding a non-trivial feasible initialization. To benchmark the proposed scheme, an SDR-based approach was also proposed, which yields an upper bound of the secure MIMO relaying problem, however, with significantly higher complexity. We compared the performance of the proposed algorithm and the benchmarking schemes using a few numerical examples. It shows that the proposed solutions yield a significantly better performance than the non-robust and null-space beamforming methods. In addition, the penalized DC algorithm often reaches performance close to the SDR-based approach.

VIII. APPENDIX A
PROOF OF LEMMA 2

From Step 2 of Algorithm 1, we obtain $(\mathbf{x}^{(n+1)}, \mathbf{S}^{(n+1)}$ is an optimal solution of the convex sub-problem (20) and $\mathbf{\Phi}_i^{(n+1)} \succeq \mathbf{0}, \mathbf{Z}_i^{(n+1)} \succeq \mathbf{0}$ for $i \in \mathcal{I}$ are the corresponding Lagrange multipliers. Since (20) is convex and strictly feasible, i.e., the Slater's constraint qualification holds, the optimal primal-dual pair must satisfy the sufficient generalized KKT conditions in (74), shown on bottom of this page, where $\mathbf{A}*$ denotes the adjoint operator of $\mathbf{A} = \sum_{i=1}^{n} x_i \mathbf{A}_i$ with $\mathbf{A}_i \in \mathbb{H}^p$ for $i = 1, \cdots, n$, i.e., $\mathbf{A} * \mathbf{Z} = [\mathrm{Tr}(\mathbf{A}_1\mathbf{Z}), \cdots, \mathrm{Tr}(\mathbf{A}_n\mathbf{Z})]^T$ for any $\mathbf{Z} \in \mathbb{H}^p$. $\mathcal{N}(\Omega, \mathbf{x})$ denotes the normal cone of Ω at $\mathbf{x}$ defined as:

$$\mathcal{N}(\Omega, \mathbf{x}) \triangleq \{\mathbf{w} \in \mathbb{C}^n | \mathbf{w}^H(\mathbf{x} - \mathbf{y}) \geq 0, \forall \mathbf{y} \in \Omega\}$$

To simply the notation, let us define $\hat{\varphi}^{(n)}(\mathbf{x}, \mathbf{S}) = \varphi(\mathbf{x}) + + \sum_{i=1}^{I} \tau^{(n)} \mathrm{Tr}(\mathbf{S}_i) = f_0(\mathbf{x}) - g_0(\mathbf{x}) + \sum_{i=1}^{I} \tau^{(n)} \mathrm{Tr}(\mathbf{S}_i)$.

First multiplying the both sides of (74a) by $(\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)})^T$ and re-arranging the consequence, we obtain

$$\begin{aligned}&\left(\nabla f_0^T(\mathbf{x}^{(n+1)}) - \nabla g_0^T(\mathbf{x}^{(n)})\right)(\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)})\\&\quad + \sum_{i=1}^{I}\left[\left(\mathcal{D}\mathcal{F}_i(\mathbf{x}^{(n+1)}) - \mathcal{D}\mathcal{G}_i(\mathbf{x}^{(n)})\right) * \mathbf{\Phi}_i\right]^T\\&\quad \times(\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)}) \geq 0 \qquad (75)\end{aligned}$$

By the assumption of convexity of $f_0(\cdot)$ and $g_0(\cdot)$, we have

$$\begin{aligned}f_0(\mathbf{x}^{(n)}) \geq\, & f_0(\mathbf{x}^{(n+1)}) + \nabla f_0^T(\mathbf{x}^{(n+1)})(\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)})\\& + \frac{\rho_f}{2}\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2 \qquad (76)\end{aligned}$$

$$\begin{aligned}g_0(\mathbf{x}^{(n+1)}) \geq\, & g_0(\mathbf{x}^{(n)}) + \nabla g_0^T(\mathbf{x}^{(n)})(\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)})\\& + \frac{\rho_g}{2}\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2, \qquad (77)\end{aligned}$$

where we recall that $\rho_f \geq 0$ and $\rho_g \geq 0$ are the convexity parameters.

Combining (76) and (77) and rearranging the consequence, we further obtain

$$\begin{aligned}&\left(\nabla f_0^T(\mathbf{x}^{(n+1)}) - \nabla g_0^T(\mathbf{x}^{(n)})\right)\left(\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)}\right)\\&\quad \leq \varphi(\mathbf{x}^{(n)}) - \varphi(\mathbf{x}^{(n+1)}) - \frac{\rho_f + \rho_g}{2}\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2.\\&\qquad (78)\end{aligned}$$

By the PSD-convexity of $\mathcal{F}_i(\cdot)$, we obtain

$$\mathcal{F}_i(\mathbf{x}^{(n)}) \succeq \mathcal{F}_i(\mathbf{x}^{(n+1)}) + \mathcal{D}\mathcal{F}_i(\mathbf{x}^{(n+1)})(\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)}), \qquad (79)$$

which further lead to

$$\begin{aligned}&\left[\mathcal{D}\mathcal{F}_i(\mathbf{x}^{(n+1)}) - \mathcal{D}\mathcal{G}_i(\mathbf{x}^{(n)})\right](\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)})\\&\quad \preceq \mathcal{F}_i(\mathbf{x}^{(n)}) - \mathcal{G}_i(\mathbf{x}^{(n)})\\&\qquad - \left[\mathcal{F}_i(\mathbf{x}^{(n+1)}) - \mathcal{G}_i(\mathbf{x}^{(n)}) - \mathcal{D}\mathcal{G}_i(\mathbf{x}^{(n)})(\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)})\right].\\&\qquad (80)\end{aligned}$$

For simplicity, let us denote the second term on the right hand side of (80) by $\mathbf{A}$. Multiplying the both sides of (80) by $\mathbf{\Phi}_i^{(n+1)} \succeq \mathbf{0}$ leads to

$$\begin{aligned}&\mathrm{Tr}\left(\mathbf{\Phi}_i^{(n+1)}\left[\mathcal{D}\mathcal{F}_i(\mathbf{x}^{(n+1)}) - \mathcal{D}\mathcal{G}_i(\mathbf{x}^{(n)})\right](\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)})\right)\\&\quad \preceq \mathrm{Tr}\left(\mathbf{\Phi}_i^{(n+1)}\left[\mathcal{F}_i(\mathbf{x}^{(n)}) - \mathcal{G}_i(\mathbf{x}^{(n)})\right]\right) + \mathrm{Tr}\left(\mathbf{\Phi}_i^{(n+1)}\mathbf{A}\right)\\&\qquad (81)\end{aligned}$$

Noting that

$$\begin{aligned}&\mathrm{Tr}\left(\mathbf{\Phi}_i^{(n+1)}\left[\mathcal{D}\mathcal{F}_i(\mathbf{x}^{(n+1)}) - \mathcal{D}\mathcal{G}_i(\mathbf{x}^{(n)})\right](\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)})\right)\\&\quad = \left[\left(\mathcal{D}\mathcal{F}_i(\mathbf{x}^{(n+1)}) - \mathcal{D}\mathcal{G}_i(\mathbf{x}^{(n)})\right) * \mathbf{\Phi}_i\right]^T(\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)})\\&\qquad (82)\end{aligned}$$

$$\mathbf{0} \in \nabla f_0(\mathbf{x}^{(n+1)}) - \nabla g_0(\mathbf{x}^{(n)}) + \sum_{i=1}^{I}\left(\left(\mathcal{D}\mathcal{F}_i(\mathbf{x}^{(n+1)}) - \mathcal{D}\mathcal{G}_i(\mathbf{x}^{(n)})\right) * \mathbf{\Phi}_i^{(n+1)}\right) + \mathcal{N}(\Omega, \mathbf{x}^{(n+1)}) \qquad (74a)$$

$$\tau^{(n)}\mathbf{I} - \mathbf{\Phi}_i^{(n+1)} - \mathbf{Z}_i^{(n+1)} = \mathbf{0},\ i \in \mathcal{I} \qquad (74b)$$

$$\mathcal{F}_i(\mathbf{x}^{(n+1)}) - \mathcal{G}_i(\mathbf{x}^{(n)}) - \mathcal{D}\mathcal{G}(\mathbf{x}^{(n)})(\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}) \preceq \mathbf{S}_i^{(n+1)},\ i \in \mathcal{I} \qquad (74c)$$

$$\mathrm{Tr}\left(\mathbf{\Phi}_i^{(n+1)}\left(\mathcal{F}_i(\mathbf{x}^{(n+1)}) - \mathcal{G}_i(\mathbf{x}^{(n)}) - \mathcal{D}\mathcal{G}(\mathbf{x}^{(n)})(\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}) - \mathbf{S}_i^{(n+1)}\right)\right) = 0,\ i \in \mathcal{I} \qquad (74d)$$

$$\mathbf{x}^{(n+1)} \in \Omega,\ \mathbf{S}_i^{(n+1)} \succeq \mathbf{0},\ \mathbf{\Phi}_i^{(n+1)} \succeq \mathbf{0},\ \mathbf{Z}_i^{(n+1)} \succeq \mathbf{0},\ \mathrm{Tr}(\mathbf{S}_i^{(n+1)}\mathbf{Z}_i^{(n+1)}) = 0,\ i \in \mathcal{I} \qquad (74e)$$

Substituting the results of (81), (82) and (74d) into (80), we have

$$\begin{aligned}&\left[\left(\mathcal{DF}_i(\mathbf{x}^{(n+1)})-\mathcal{DG}_i(\mathbf{x}^{(n)})\right)*\mathbf{\Phi}_i\right]^T(\mathbf{x}^{(n)}-\mathbf{x}^{(n+1)})\\&\quad\le\operatorname{Tr}\left(\mathbf{\Phi}_i^{(n+1)}\left[\mathcal{F}_i(\mathbf{x}^{(n)})-\mathcal{G}_i(\mathbf{x}^{(n)})\right]\right)\\&\quad-\operatorname{Tr}\left(\mathbf{\Phi}_i^{(n+1)}\mathbf{S}_i^{(n+1)}\right).\end{aligned}\tag{83}$$

Observing that $\mathcal{F}_i(\mathbf{x}^{(n)})-\mathcal{G}_i(\mathbf{x}^{(n)})\preceq\mathbf{S}_i^{(n)}$ and $\tau^{(n)}\mathbf{I}\succeq\mathbf{\Phi}_i^{(n+1)}$ [c.f., (74b)], (83) can further be derived as

$$\begin{aligned}&\left[\left(\mathcal{DF}_i(\mathbf{x}^{(n+1)})-\mathcal{DG}_i(\mathbf{x}^{(n)})\right)*\mathbf{\Phi}_i\right]^T(\mathbf{x}^{(n)}-\mathbf{x}^{(n+1)})\\&\quad\le\operatorname{Tr}\left(\mathbf{\Phi}_i^{(n+1)}\left(\mathbf{S}_i^{(n)}-\mathbf{S}_i^{(n+1)}\right)\right)\\&\quad\le\tau^{(n)}\operatorname{Tr}(\mathbf{S}_i^{(n)})-\tau^{(n)}\operatorname{Tr}(\mathbf{S}_i^{(n+1)}).\end{aligned}\tag{84}$$

Combining (75), (78) and (84), we have reached:

$$\begin{aligned}&\hat{\varphi}^{(n)}(\mathbf{x}^{(n)},\mathbf{S}^{(n)})-\hat{\varphi}^{(n)}(\mathbf{x}^{(n+1)},\mathbf{S}^{(n+1)})\\&\quad\ge\frac{\rho_f+\rho_g}{2}\|\mathbf{x}^{(n+1)}-\mathbf{x}^{(n)}\|^2.\end{aligned}\tag{85}$$

The above inequality is indeed (37), which therefore proves the item 1). If either f_0 or g_0 is strongly convex, i.e., $\rho_f+\rho_g>0$, then the statement in item 2) follows directly from the above inequality, i.e., for $\Delta\mathbf{x}^{(n)}=\mathbf{x}^{(n+1)}-\mathbf{x}^{(n)}\neq\mathbf{0}$, $\hat{\varphi}^{(n)}(\mathbf{x}^{(n+1)},\mathbf{S}^{(n+1)})<\hat{\varphi}^{(n)}(\mathbf{x}^{(n)},\mathbf{S}^{(n)})$.

IX. APPENDIX B
PROOF OF THEOREM 1

We first prove scenario 1). If Algorithm 1 terminates after a finite number of $\check{n}$ iterations, it follows from the termination criterion that $\mathbf{x}^{(\check{n}+1)}=\mathbf{x}^{(\check{n})}$ and $\mathbf{S}_i^{(\check{n}+1)}=\mathbf{0}$ for all i, i.e., $\check{\mathbf{x}}$ is a feasible solution to (20). Letting $n=\check{n}$ and substituting the above relations into the generalized KKT conditions (74), we obtain (86), shown on bottom of this page. A careful examination reveals the equivalence between (86) and the KKT conditions of (20). Therefore, it is proved that $(\mathbf{x}^{(\check{n})},\{\mathbf{\Phi}_i^{(\check{n})}\}$ is a KKT point of (16), where $\mathbf{x}^{(\check{n})}$ is called a stationary point of (20) and $\{\mathbf{\Phi}_i^{(\check{n})}\}$ are the corresponding Lagrange multipliers

We now proceed to prove scenario 2). The key ingredients of the proof are to show that any limit point of $\{\mathbf{x}^{(n)}\}$, say, $\bar{\mathbf{x}}$, is feasible to (20) and the sequence of dual variable $\{\mathbf{\Phi}_i^{(n)}\}$ is bounded such that there exists limit points $\bar{\mathbf{\Phi}}_i$ of $\{\mathbf{\Phi}_i^{(n)}\}$. Then we show that any primal-dual pair of the limit point $(\bar{\mathbf{x}},\{\bar{\mathbf{\Phi}}_i\})$ satisfies the KKT conditions of (20).

To prove that any limit point $\bar{\mathbf{x}}$ is a feasible point of (16), we will need to rely on the following claims, whose proof can be found in Appendices X and XI, respectively:

Claim 2 There exists a finite iteration index $\tilde{n}$ such that

$$\tau^{(n)}=\tau^{(\tilde{n})},\quad\forall n\ge\tilde{n}.\tag{87}$$

Claim 3 The sequence of intermediate solutions $\{\mathbf{x}^{(n)}\}$ satisfies

$$\lim_{n\to\infty}\|\mathbf{x}^{(n+1)}-\mathbf{x}^{(n)}\|=0.\tag{88}$$

As indicated by the updating rule (30), we have

$$\tau_i^{(\tilde{n})}\ge\lambda_{\max}[\mathbf{\Phi}_i^{(\tilde{n}+1)}]+\delta_1,\ i\in\mathcal{I}\tag{89}$$

or equivalently, $\tau^{(n)}\mathbf{I}\succeq\mathbf{\Phi}_i^{(n+1)}+\delta_1\mathbf{I}$ for all $n\ge\tilde{n}$. Then in view of the complementary slackness (74b), it straightforwardly follows that $\mathbf{Z}_i^{(n+1)}\succ\mathbf{0}$. By (74e), we obtain $\mathbf{S}_i^{(n+1)}=\mathbf{0}$ for all $n\ge\check{n}$, which means that $\mathbf{x}^{(n)}$ is a feasible point of (20) for all $n\ge\tilde{n}$. Without loss of generality, considering a subsequence $\{\mathbf{x}^{(n_j)}\}$ of $\{\mathbf{x}^{(n)}\}$, its limit point $\lim_{j\to\infty}\mathbf{x}^{(n_j)}=\bar{\mathbf{x}}$ is feasible to (16). Furthermore, (89) implies that the subsequence $\{\mathbf{\Phi}_i^{(n_j)}\}$ is bounded, and therefore we can assume that

$$\lim_{j\to\infty}\mathbf{\Phi}_i^{(n_j)}=\bar{\mathbf{\Phi}}_i,\ i\in\mathcal{I}.\tag{90}$$

Now what remains to show is that any primal-dual pair of the limit point $(\bar{\mathbf{x}},\{\bar{\mathbf{\Phi}}_i\})$ a KKT stationary point of (20). Let us replace n with n_j in (74) and let $j\to\infty$. By noting that $\mathbf{x}^{(n_j)}$ and $\mathbf{x}^{(n_j+1)}$ are asymptotically close as indicated by Claim 3, we obtain

$$\mathbf{0}\in\nabla f_0(\bar{\mathbf{x}})-\nabla g_0(\bar{\mathbf{x}})$$
$$+\sum_{i=1}^{I}\left(\bar{\mathbf{\Phi}}_i*(\mathcal{DF}_i(\bar{\mathbf{x}})-\mathcal{DG}_i(\bar{\mathbf{x}}))\right)+\mathcal{N}(\Omega,\bar{\mathbf{x}})\tag{91a}$$
$$\mathcal{F}_i(\bar{\mathbf{x}})-\mathcal{G}_i(\bar{\mathbf{x}})\preceq\mathbf{0},\ \bar{\mathbf{\Phi}}_i\succeq\mathbf{0},\ i\in\mathcal{I}\tag{91b}$$
$$\operatorname{Tr}\left((\mathcal{F}_i(\bar{\mathbf{x}})-\mathcal{G}_i(\bar{\mathbf{x}}))\,\bar{\mathbf{\Phi}}_i\right)=0,\ i\in\mathcal{I}\tag{91c}$$
$$\bar{\mathbf{x}}\in\Omega\tag{91d}$$

which is exactly the KKT conditions of the PSD DC problem (20). Noting the boundness of $\{\mathbf{x}^{(n)}\}$ assumed in A.2), it readily follows that there exists at least one limit point of $\{\mathbf{x}^{(n)}\}$ and by (91), any limit point of $\left\{\mathbf{x}^{(n)}\right\}$ is a KKT stationary point of (20).

X. APPENDIX C
PROOF OF CLAIM 2

We argue by contradiction. Assume the contrary, i.e., $\lim_{n\to\infty}\tau^{(n)}=+\infty$. From the updating rule (30), it follows,

$$\mathbf{0}\in\nabla f_0(\mathbf{x}^{(\check{n})})-\nabla g_0(\mathbf{x}^{(\check{n})})+\sum_{i=1}^{I}\left(\left(\mathcal{DF}_i(\mathbf{x}^{(\check{n})})-\mathcal{DG}_i(\mathbf{x}^{(\check{n})})\right)*\mathbf{\Phi}_i^{(\check{n})}\right)+\mathcal{N}(\Omega,\mathbf{x}^{(\check{n})})\tag{86a}$$
$$\mathcal{F}_i(\mathbf{x}^{(\check{n})})-\mathcal{G}_i(\mathbf{x}^{(\check{n})})\preceq\mathbf{0},\ i\in\mathcal{I}\tag{86b}$$
$$\operatorname{Tr}\left(\mathbf{\Phi}_i^{(\check{n}+1)}\left(\mathcal{F}_i(\mathbf{x}^{(\check{n})})-\mathcal{G}_i(\mathbf{x}^{(\check{n})})\right)\right)=0,\ i\in\mathcal{I}\tag{86c}$$
$$\mathbf{x}^{(\check{n})}\in\Omega,\ \mathbf{\Phi}_i^{(\check{n}+1)}\succeq\mathbf{0}.\tag{86d}$$

without loss of generality, that there exists infinitely many indices j such that

$$\tau^{(n_j)} < \lambda_{\max}\left[\sum_{i=1}^{I} \mathbf{\Phi}_i^{(n_j+1)}\right] + \delta_1 \tag{92}$$

and

$$\tau^{(n_j)} < \|\mathbf{x}^{(n_j+1)} - \mathbf{x}^{(n_j)}\|^{-1}. \tag{93}$$

By possibly restricting to a subsequence of $\{n_j\}$, without loss of generality, we can further assume that there exists at least some $i \in \mathcal{S}_{\mathcal{I}}$, where $\mathcal{S}_{\mathcal{I}}$ denotes a subset of $\mathcal{I}$, i.e., $\mathcal{S}_{\mathcal{I}} \subseteq \mathcal{I}$ such that

$$\begin{aligned}&\lim_{j\to\infty} \lambda_{\max}[\mathbf{\Phi}_i^{(n_j+1)}] = +\infty\\ &\Leftrightarrow \lim_{j\to\infty} \|\mathbf{\Phi}_i^{(n_j+1)}\|_F = +\infty,\ i \in \mathcal{S}_{\mathcal{I}}\end{aligned} \tag{94}$$

and

$$\lim_{j\to\infty} \|\mathbf{x}^{(n_j+1)} - \mathbf{x}^{(n_j)}\| = 0. \tag{95}$$

Let $\lim_{j\to\infty} \mathbf{x}^{n_j} = \bar{\mathbf{x}}$, and then we will show that

$$\boldsymbol{\mathcal{F}}_i(\bar{\mathbf{x}}) - \boldsymbol{\mathcal{G}}_i(\bar{\mathbf{x}}) \not\prec \mathbf{0},\ i \in \mathcal{S}_{\mathcal{I}}. \tag{96}$$

Again we show by contradiction. If we assume that $\boldsymbol{\mathcal{F}}_i(\bar{\mathbf{x}}) - \boldsymbol{\mathcal{G}}_i(\bar{\mathbf{x}}) \prec \mathbf{0}$, then we must have, for sufficiently large j,

$$\begin{aligned}&\boldsymbol{\mathcal{F}}_i(\mathbf{x}^{(n_j+1)}) - \boldsymbol{\mathcal{G}}_i(\mathbf{x}^{(n_j)})\\ &\quad -\boldsymbol{\mathcal{DG}}(\mathbf{x}^{(n_j)})(\mathbf{x}^{(n_j+1)} - \mathbf{x}^{(n_j)}) - \mathbf{S}_i^{(n_j+1)} \prec \mathbf{0}.\end{aligned} \tag{97}$$

This is due to (74b) and subsequently

$$\begin{aligned}&\lim_{j\to\infty}\Big(\boldsymbol{\mathcal{F}}_i(\mathbf{x}^{(n_j+1)}) - \boldsymbol{\mathcal{G}}_i(\mathbf{x}^{(n_j)}) - \boldsymbol{\mathcal{DG}}(\mathbf{x}^{(n_j)})(\mathbf{x}^{(n_j+1)} - \mathbf{x}^{(n_j)})\\ &\quad -\mathbf{S}_i^{(n_j+1)}\Big) = \boldsymbol{\mathcal{F}}_i(\bar{\mathbf{x}}) - \boldsymbol{\mathcal{G}}_i(\bar{\mathbf{x}}).\end{aligned} \tag{98}$$

By the complementary slackness condition (74d), it readily follows that when j becomes sufficiently large, $\mathbf{\Phi}_i^{(n_j+1)} = \mathbf{0}$ for $i \in \mathcal{S}_{\mathcal{I}}$, which contradicts the previous result of (92). Therefore, we must have $\boldsymbol{\mathcal{F}}_i(\bar{\mathbf{x}}) - \boldsymbol{\mathcal{G}}_i(\bar{\mathbf{x}}) \not\prec \mathbf{0}$ for $i \in \mathcal{S}_{\mathcal{I}}$.

Now let us assume, without loss of generality, that

$$\lim_{j\to\infty} \frac{\mathbf{\Phi}_i^{(n_j+1)}}{\sum_{i=1}^{I} \|\mathbf{\Phi}_i^{(n_j+1)}\|_F} = \hat{\mathbf{\Phi}}_i \succeq \mathbf{0}. \tag{99}$$

and it is easy to observe that $\hat{\mathbf{\Phi}}_i = \mathbf{0}$ for $i \in \mathcal{I}\backslash\mathcal{S}_{\mathcal{I}}$ and $\hat{\mathbf{\Phi}}_i \neq \mathbf{0}$ for $i \in \mathcal{S}_{\mathcal{I}}$. We now replace n with n_j in (39). Dividing the both sides of (74a) by $\sum_{i=1}^{I} \|\mathbf{\Phi}_i^{(n_j+1)}\|_F$, taking the limit as $j \to \infty$ and using the result of (97), we obtain

$$\mathbf{0} \in \sum_{i\in\mathcal{S}_{\mathcal{I}}} (\boldsymbol{\mathcal{DF}}_i(\bar{\mathbf{x}}) - \boldsymbol{\mathcal{DG}}_i(\bar{\mathbf{x}})) * \hat{\mathbf{\Phi}}_i + \mathcal{N}(\Omega, \bar{\mathbf{x}}). \tag{100}$$

Multiplying the both sides of the above by $(\mathbf{y} - \mathbf{x}), \mathbf{y} \in \Omega$ yields

$$\sum_{i\in\mathcal{S}_{\mathcal{I}}} \mathrm{Tr}\left(\hat{\mathbf{\Phi}}_i\,(\boldsymbol{\mathcal{DF}}_i(\bar{\mathbf{x}}) - \boldsymbol{\mathcal{DG}}_i(\bar{\mathbf{x}}))\,(\mathbf{y} - \bar{\mathbf{x}})\right) \geq 0. \tag{101}$$

However, the MFCQ in A.1) indicates that there exists some feasible direction $\mathbf{h} \in \texttt{cone}(\Omega - \bar{\mathbf{x}})$ such that

$$(\boldsymbol{\mathcal{DF}}_i(\bar{\mathbf{x}}) - \boldsymbol{\mathcal{DG}}_i(\bar{\mathbf{x}}))\,\mathbf{h} \prec \mathbf{0},\ \forall i \in \mathcal{U}(\bar{\mathbf{x}}), \tag{102}$$

where we recall that $\mathcal{U}(\bar{\mathbf{x}})$ is the set of *active* constraints at $\bar{\mathbf{x}}$:

$$\mathcal{U}(\bar{\mathbf{x}}) \triangleq \left\{i \in \mathcal{I} \middle| \boldsymbol{\mathcal{F}}_i(\bar{\mathbf{x}}) - \boldsymbol{\mathcal{G}}_i(\bar{\mathbf{x}}) \not\prec \mathbf{0}\right\}. \tag{103}$$

Considering $\hat{\mathbf{\Phi}}_i \succeq \mathbf{0}$, it is obvious (101) contradicts the MFCQ in A.1).

Now we can assume that there exists an finite index $\tilde{n}$ such that

$$\tau^{(n)} = \tau^{(\tilde{n})},\ \forall n \geq \tilde{n}. \tag{104}$$

XI. APPENDIX D
PROOF OF CLAIM 3

Following directly from Lemma 2, we have

$$\begin{aligned}&\hat{\varphi}^{(n)}(\mathbf{x}^{(n)}, \mathbf{S}^{(n)}) - \hat{\varphi}^{(n)}(\mathbf{x}^{(n+1)}, \mathbf{S}^{(n+1)})\\ &\geq \frac{\rho_f + \rho_g}{2}\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2.\end{aligned} \tag{105}$$

Then we evaluate the summation of (105) over n from 0 to $\tilde{n}$ and we obtain

$$\begin{aligned}&\sum_{n=0}^{\tilde{n}} \frac{\rho_f + \rho_g}{2}\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2\\ &\leq \sum_{n=0}^{\tilde{n}}\left(\hat{\varphi}^{(n)}(\mathbf{x}^{(n)}, \mathbf{S}^{(n)}) - \hat{\varphi}^{(n)}(\mathbf{x}^{(n+1)}, \mathbf{S}^{(n+1)})\right)\\ &\quad + \hat{\varphi}^{(\tilde{n})}(\mathbf{x}^{(\tilde{n})}, \mathbf{S}^{(\tilde{n})}) - \hat{\varphi}^{(\tilde{n})}(\mathbf{x}^{(\tilde{n}+1)}, \mathbf{S}^{(\tilde{n}+1)})\end{aligned} \tag{106}$$

where $\tilde{n}$ is a finite index as defined in (104), such that $\tau^{(n)} = \tau^{(\tilde{n})}$, $\forall n \geq \tilde{n}$, i.e., the value of τ remains constant for all indices $n \geq \tilde{n}$. By A.2), $\varphi(\mathbf{x})$ is bounded from below and hence $\hat{\varphi}^{(n)}(\cdot)$ is also lower-bounded. Taking the limit as $\tilde{n} \to \infty$ on both sides of (106), we obtain

$$\begin{aligned}&\sum_{n=0}^{\infty} \frac{\rho_f + \rho_g}{2}\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2\\ &\leq \sum_{n=0}^{\tilde{n}}\left(\hat{\varphi}^{(n)}(\mathbf{x}^{(n)}, \mathbf{S}^{(n)}) - \hat{\varphi}^{(n)}(\mathbf{x}^{(n+1)}, \mathbf{S}^{(n+1)})\right)\\ &\quad + \hat{\varphi}^{(\tilde{n})}(\mathbf{x}^{(\tilde{n})}, \mathbf{S}^{(\tilde{n})}) - \hat{\varphi}^{(\infty)}(\bar{\mathbf{x}}, \bar{\mathbf{S}}).\end{aligned} \tag{107}$$

The right hand side of the above inequality must be of finite value. Therefore, it must hold that

$$\lim_{n\to\infty} \|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\| = 0. \tag{108}$$

XII. APPENDIX E
PROOF OF PROPOSITION 2

Firstly, expanding all the quadratic terms of $\mathbf{W}$ in (66a) and (66b), and invoking the identities $\mathrm{Tr}\left(\mathbf{A}^H\mathbf{BCD}^H\right) = \mathrm{vec}\,(\mathbf{A})^H\left(\mathbf{D}^T \otimes \mathbf{B}\right)\mathrm{vec}\,(\mathbf{C})$ and $\mathrm{Tr}\left(\mathbf{A}^H\mathbf{BA}\right) = \mathrm{vec}\,(\mathbf{A})^H\,(\mathbf{I} \otimes \mathbf{B})\,\mathrm{vec}\,(\mathbf{A})$, (66a) and (66b) can be recast as (71a) and (71b), respectively, with $\mathbf{X} = \mathbf{w}\mathbf{w}^H$.

Next, we transform (66c). Recall that $\boldsymbol{\Theta}(\mathbf{W}, \boldsymbol{\Psi}) = \sigma_{\mathrm{S}}^2 \mathbf{W}\mathbf{h}_1\mathbf{h}_1^H\mathbf{W}^H - \gamma\sigma_{\mathrm{R}}^2\mathbf{W}\mathbf{W}^H - \gamma\boldsymbol{\Psi}$, and its first term on the right hand side is equivalent to

$$\sigma_{\mathrm{S}}^2 \mathbf{W}\mathbf{h}_1\mathbf{h}_1^H\mathbf{W}^H = \sigma_{\mathrm{S}}^2(\mathbf{h}_1^T \otimes \mathbf{I}_{N_{\mathrm{R}}})\mathbf{w}\mathbf{w}^H(\mathbf{h}_1^T \otimes \mathbf{I}_{N_{\mathrm{R}}})^H, \tag{109}$$

by using $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$. To transform the second term, we express $\mathbf{W} = \left[\boldsymbol{w}_1, \cdots, \boldsymbol{w}_l, \cdots, \boldsymbol{w}_{N_{\mathrm{R}}}\right]$, where $\boldsymbol{w}_l$ denotes the l^{th} column of $\mathbf{W}$. Then $\mathbf{W}\mathbf{W}^H$ can be equivalently expressed as

$$\mathbf{W}\mathbf{W}^H = \sum_{l=1}^{N_{\mathrm{R}}} \boldsymbol{w}_l\boldsymbol{w}_l^H \tag{110}$$

By establishing the connection between $\boldsymbol{w}_l$ and $\mathbf{w}$ by $\boldsymbol{w}_l = \mathbf{E}_l\mathbf{w}$, (110) can further be written as

$$\mathbf{W}\mathbf{W}^H = \sum_{l=1}^{N_{\mathrm{R}}} \mathbf{E}_l\mathbf{w}\mathbf{w}^H\mathbf{E}_l^H. \tag{111}$$

Using (109) and (111), $\boldsymbol{\Theta}(\mathbf{W}, \boldsymbol{\Psi})$ is equivalent to

$$\boldsymbol{\Theta}(\mathbf{W}, \boldsymbol{\Psi}) = \sigma_{\mathrm{S}}^2\mathbf{H}_1\mathbf{w}\mathbf{w}^H\mathbf{H}_1^H - \gamma\sigma_{\mathrm{R}}^2\sum_{l=1}^{N_{\mathrm{R}}}\mathbf{E}_l\mathbf{w}\mathbf{w}^H\mathbf{E}_l^H - \gamma\boldsymbol{\Psi}. \tag{112}$$

Invoking $\mathbf{X} = \mathbf{w}\mathbf{w}^H$ and $\mathrm{Rank}(\mathbf{X}) = 1$, (66) is readily re-expressed as (71).

XIII. APPENDIX F
PROOF OF PROPOSITION 3

We prove the rank-one optimality of the solution to (73) by examining its Karush-Kuhn-Tucker (KKT) conditions. Let y_1, y_2 and $\mathbf{Y}_k$ denote the Lagrange multipliers associated with (73b)–(73d), respective, and let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ denote the Lagrange multipliers associated with $\mathbf{X} \succeq \mathbf{0}$ and $\boldsymbol{\Psi} \succeq \mathbf{0}$, respectively. The Lagrangian function of (73) can then be written as

$$\begin{aligned}\mathcal{L} = {} & \mathrm{Tr}\,(\mathbf{Q}_2\mathbf{X}) + y_1(\mathrm{Tr}(\mathbf{Q}_2\mathbf{X}) + \mathrm{Tr}(\boldsymbol{\Psi}))\\ & - y_2\left(\mathrm{Tr}(\mathbf{Q}_0\mathbf{X}) - \tau^*\,\mathrm{Tr}(\mathbf{Q}_1\mathbf{X}) - \tau^*\,\mathrm{Tr}(\mathbf{h}_2\mathbf{h}_2^H\boldsymbol{\Psi})\right)\\ & + \sum_{k=1}^{K}\mathrm{Tr}\left(\mathbf{P}_k^H\mathbf{Q}_3(\mathbf{X}, \boldsymbol{\Psi})\mathbf{P}_k\mathbf{Y}_k\right)\\ & - \mathrm{Tr}(\mathbf{X}\mathbf{Z}_1) - \mathrm{Tr}(\boldsymbol{\Psi}\mathbf{Z}_2),\end{aligned} \tag{113}$$

where we have neglected the terms, which are independent of $\mathbf{X}$ and $\boldsymbol{\Psi}$. Now we exploit the first-order KKT conditions with respect to $\mathbf{X}$ and $\boldsymbol{\Psi}$, which can be given by

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\mathbf{X}} = {} & \mathbf{Q}_2 + y_1\mathbf{Q}_2 - y_2\mathbf{Q}_0 + y_2\tau^*\mathbf{Q}_1\\ & + \sum_{k=1}^{K}\sigma_{\mathrm{S}}^2\mathbf{H}_1^H\mathbf{P}_k\mathbf{Y}_k\mathbf{P}_k^H\mathbf{H}_1\\ & - \gamma\sigma_{\mathrm{R}}^2\sum_{k=1}^{K}\sum_{l=1}^{N_{\mathrm{R}}}\mathbf{E}_l^H\mathbf{P}_k\mathbf{Y}_k\mathbf{P}_k^H\mathbf{E}_l - \mathbf{Z}_1 = \mathbf{0}\end{aligned} \tag{114}$$

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{\Psi}} = y_1\mathbf{I} + y_2\tau^*\mathbf{h}_2\mathbf{h}_2^H - \sum_{k=1}^{K}\gamma\mathbf{P}_k\mathbf{Y}_k\mathbf{P}_k^H - \mathbf{Z}_2 = \mathbf{0} \tag{115}$$

Re-arranging (115) and using the associativity of the Kronecker product, we obtain

$$\begin{aligned}\mathbf{I}\otimes\mathbf{Z}_2 = {} & y_1\mathbf{I} + y_2\tau^*\mathbf{I}\otimes(\mathbf{h}_2\mathbf{h}_2^H)\\ & - \mathbf{I}\otimes\sum_{k=1}^{K}\gamma\mathbf{P}_k\mathbf{Y}_k\mathbf{P}_k^H.\end{aligned} \tag{116}$$

A simple calculation reveals that the right hand side of (116) is equivalent to

$$\begin{aligned}\mathbf{I}\otimes\mathbf{Z}_2 = {} & y_1\mathbf{I} + y_2\tau^*\mathbf{Q}_1/\sigma_{\mathrm{R}}^2\\ & - \gamma\sum_{k=1}^{K}\sum_{l=1}^{N_{\mathrm{R}}}\mathbf{E}_l^H\mathbf{P}_k\mathbf{Y}_k\mathbf{P}_k^H\mathbf{E}_l.\end{aligned} \tag{117}$$

Substituting the above relation into (114), we further obtain

$$\begin{aligned}& \underbrace{\mathbf{Q}_2 + y_1\mathbf{Q}_2 + \sigma_{\mathrm{R}}^2\mathbf{Z}_2 + \sum_{k=1}^{K}\sigma_{\mathrm{S}}^2\mathbf{H}_1^H\mathbf{P}_k\mathbf{Y}_k\mathbf{P}_k^H\mathbf{H}_1}_{\triangleq\boldsymbol{\Theta}}\\ & \qquad - y_2\mathbf{Q}_0 = \mathbf{Z}_1.\end{aligned} \tag{118}$$

Since $\mathbf{Q}_2 \succ \mathbf{0}$, $\boldsymbol{\Theta}$ must be a positive definite matrix, which has full rank, i.e., $\mathrm{Rank}(\boldsymbol{\Theta}) = N_{\mathrm{R}}^2$. It is further implied by (118) that

$$\mathrm{Rank}(\mathbf{Z}_1) \geq \mathrm{Rank}(\boldsymbol{\Theta}) - \mathrm{Rank}(\mathbf{Q}_0), \tag{119}$$

where $\mathrm{Rank}(\mathbf{Q}_0) = 1$. Then it is clear that the rank of $\mathbf{Z}_1$ is either N_{R}^2 or $N_{\mathrm{R}}^2 - 1$. If $\mathrm{Rank}(\mathbf{Z}_1) = N_{\mathrm{R}}^2$, we must have $\mathbf{X} = \mathbf{0}$ due to the complementary slackness condition $\mathrm{Tr}(\mathbf{X}\mathbf{Z}_1) = 0$. However, it is obvious that $\mathbf{X} = \mathbf{0}$ is not the optimal solution. Then the rank of $\mathbf{Z}_1$ must be $N_{\mathrm{R}}^2 - 1$ and in this case, $\mathbf{X}$ must lie in the nullspace of $\mathbf{Z}_1$, whose dimension is one. Therefore, $\mathbf{X}$ must be of rank one.

REFERENCES

[1] "Mobile cyber threats: Kaspersky Lab & INTERPOL joint report, " INTERPOL Kaspersky Lab., Tech. Rep., Oct. 2014.

[2] C. Timberg. (Dec. 2014). *German Researchers Discover a Flaw That Could Let Anyone Listen to Your Cell Calls*, The Switch, Washington, DC, USA. [Online]. Available: https://www.washingtonpost.com/news/the-switch/wp/2014/12/18/german-researchers-discover-a-flaw-that-could-let-anyone-listen-to-your-cell-calls-and-read-your-texts/

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

[4] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Information theoretic security," *Found. Trends Commun. Inf. Theory*, vol. 5, nos. 4–5, pp. 355–580, 2008.

[5] Y.-W. P. Hong, P.-C. Lan, and C.-C. J. Kuo, "Enhancing physical-layer secrecy in multiantenna wireless systems: An overview of signal processing approaches," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 29–40, Sep. 2013.

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

[7] C. Xing, S. Ma, and Y.-C. Wu, "Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2273–2283, Apr. 2010.

[8] Y. Rong, X. Tang, and Y. Hua, "A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4837–4851, Dec. 2009.

[9] Z. Ho and E. Jorswieck, "Signal leakage neutralisation in instantaneous non-regenerative relaying networks under channel uncertainty," *IET Commun.*, vol. 8, no. 8, pp. 1285–1295, May 2014.

[10] L. Lai and H. El Gamal, "The relay–eavesdropper channel: Cooperation for secrecy," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4005–4019, Sep. 2008.

[11] H.-M. Wang and X.-G. Xia, "Enhancing wireless secrecy via cooperation: Signal design and optimization," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 47–53, Dec. 2015.

[12] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Improving wireless physical layer security via cooperating relays," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1875–1888, Mar. 2010.

[13] J. Li, A. P. Petropulu, and S. Weber, "On cooperative relaying schemes for wireless physical layer security," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4985–4997, Oct. 2011.

[14] H. Deng, H. M. Wang, W. Guo, and W. Wang, "Secrecy transmission with a helper: To relay or to jam," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 2, pp. 293–307, Feb. 2015.

[15] Y. Yang, Q. Li, W.-K. Ma, J. Ge, and P. C. Ching, "Cooperative secure beamforming for AF relay networks with multiple eavesdroppers," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 35–38, Jan. 2013.

[16] H.-M. Wang, F. Liu, and X.-G. Xia, "Joint source-relay precoding and power allocation for secure amplify-and-forward MIMO relay networks," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 8, pp. 1240–1250, Aug. 2014.

[17] C. Jeong, I.-M. Kim, and D. Kim, "Joint secure beamforming design at the source and the relay for an amplify-and-forward MIMO untrusted relay system," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 310–325, Jan. 2012.

[18] S. Vishwakarma and A. Chockalingam, "Amplify-and-forward relay beamforming for secrecy with cooperative jamming and imperfect CSI," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungry, Jun. 2013, pp. 3047–3052.

[19] C. Zhang, H. Gao, H. Liu, and T. Lv, "Robust beamforming and jamming for secure AF relay networks with multiple eavesdroppers," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Baltimore, MD, USA, Oct. 2014, pp. 495–500.

[20] X. Wang, K. Wang, and X.-D. Zhang, "Secure relay beamforming with imperfect channel side information," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2140–2155, Jun. 2013.

[21] Q. Li, Y. Yang, W. K. Ma, M. Lin, J. Ge, and J. Lin, "Robust cooperative beamforming and artificial noise design for physical-layer secrecy in AF multi-antenna multi-relay networks," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 206–220, Jan. 2015.

[22] K. Jayasinghe, P. Jayasinghe, N. Rajatheva, and M. Latva-Aho, "Secure beamforming design for physical layer network coding based MIMO two-way relaying," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1270–1273, Jul. 2014.

[23] M. Zhang, J. Huang, H. Yu, H. Luo, and W. Chen, "QoS-based source and relay secure optimization design with presence of channel uncertainty," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1544–1547, Aug. 2013.

[24] Z. Chu, K. Cumanan, M. Xu, and Z. Ding, "Robust secrecy rate optimisations for multiuser multiple-input-single-output channel with device-to-device communications," *IET Commun.*, vol. 9, no. 3, pp. 396–403, Feb. 2015.

[25] C. Wang and H.-M. Wang, "Robust joint beamforming and jamming for secure AF networks: Low-complexity design," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 2192–2198, May 2015.

[26] H.-M. Wang, F. Liu, and M. Yang, "Joint cooperative beamforming, jamming, and power allocation to secure AF relay systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4893–4898, Oct. 2015.

[27] J. Yang, B. Champagne, Q. Li, and L. Hanzo, "Secure MIMO AF relaying design: An intercept probability constrained approach," in *Proc. IEEE Global Commun. Conf. (Globecom)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[28] 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.

[29] B. K. Sriperumbudur and G. R. Lanckriet, "On the convergence of the concave-convex procedure," in *Proc. 22nd Adv. Neural Inf. Process. Syst.*, 2009, pp. 1759–1767.

[30] A. Pascual-Iserte, D. P. Palomar, A. I. Perez-Neira, and M. A. Lagunas, "A robust maximin approach for MIMO communications with imperfect channel state information based on convex optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 346–360, Jan. 2006.

[31] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on LTE Device to Device Proximity Services; Radio Aspects (Release 12)*, document 3GPP TR 36.843 V12.0.1, Mar. 2014.

[32] A. Goldsmith, *Wireless Communications*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, Aug. 2005.

[33] R. Horst and N. V. Thoai, "DC programming: Overview," *J. Optim. Theory Appl.*, vol. 103, no. 1, pp. 1–43, Oct. 1999.

[34] U. Rashid, H. D. Tuan, H. H. Kha, and H. H. Nguyen, "Joint optimization of source precoding and relay beamforming in wireless MIMO relay networks," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 488–499, Feb. 2014.

[35] S.-P. Han and O. L. Mangasarian, "Exact penalty functions in nonlinear programming," *Math. Program.*, vol. 17, no. 1, pp. 251–269, Dec. 1979.

[36] G. Di Pillo and L. Grippo, "Exact penalty functions in constrained optimization," *SIAM J. Control Optim.*, vol. 26, no. 6, pp. 1333–1360, Nov. 1989.

[37] O. L. Mangasarian and S. Fromovitz, "The Fritz John necessary optimality conditions in the presence of equality and inequality constraints," *J. Math. Anal. Appl.*, vol. 17, pp. 37–47, 1967.

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

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

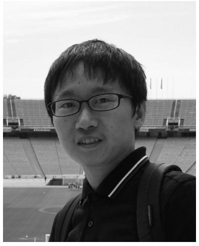
[40] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Res. Logistics Quart.*, vol. 9, nos. 3–4, pp. 181–186, Sep./Dec. 1962.

[41] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, nos. 1–4, pp. 625–653, Jan. 1999.

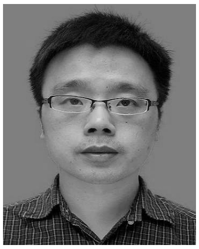
[42] MOSEK ApS. (2016). *The MOSEK Optimization Toolbox for MATLAB Manual. V7.1 (R49)*. [Online]. Available: http://docs.mosek.com/7.1/toolbox/index.html

[43] Y. Nesterov and A. Nemirovski, *Interior Point Polynomial Time Methods in Convex Programming: Theory and Applications*. Philadelphia, PA, USA: SIAM, 1994.

[44] J. Yang, B. Champagne, Y. Zou, and L. Hanzo, "Joint optimization of transceiver matrices for MIMO-aided multiuser AF relay networks: Improving the QoS in the presence of CSI errors," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1434–1451, Mar. 2016.

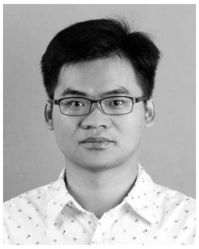


JIAXIN YANG (S'11) received the B.Eng. degree in information engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009, and the M.E.Sc. degree in electrical and computer engineering from the University of Western Ontario, London, ON, Canada, in 2011. Since 2012, he has been with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada, where he is currently pursuing the Ph.D. degree. He has been a Wireless System Research Intern with InterDigital since 2015. His research interests include optimization theory, statistical signal processing, detection and estimation, and the applications thereof in wireless communications such as MIMO systems, co-operative communications, and physical-layer security. He was a recipient of several awards and scholarships, including the Best Paper Award of the 27th IEEE International Symposium on Personal, the Indoor and Mobile Radio Communications, the McGill Engineering Doctoral Award, the Graduate Excellence Fellowship, the International Differential Tuition FeeWaivers, the FRQNT International Internship Scholarship, the PERSWADE Ph.D. Scholarship, the Graduate Research Enhancement and Travel Awards, and the Graduate Research Mobility Awards.



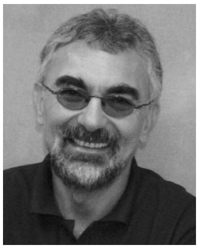
QIANG LI (M'13) received the B.Eng. and M.Phil. degrees in communication and information engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2005 and 2008, respectively, and the Ph.D. degree in electronic engineering from the Chinese University of Hong Kong (CUHK), Hong Kong, in 2012. From 2011 to 2012, he was a Visiting Scholar with the University of Minnesota, Twin Cities, Minneapolis, MN, USA. From 2012 to 2013, he was a Research Associate with the Department of Electronic Engineering and the Department of Systems Engineering and Engineering Management, CUHK. Since 2013, he has been with the School of Communication and Information Engineering, UESTC, where he is currently an Associate Professor. His research interests include convex optimization and its applications in signal processing with an emphasis on the physical-layer security and full-duplex communications. He was a recipient of the First Prize Paper Award in the IEEE Signal Processing Society Postgraduate Forum Hong Kong Chapter in 2010, and a co-recipient of the Best Paper Award of the IEEE PIMRC 2016.

YUNLONG CAI (S'07–M'10–SM'16) received the B.S. degree in computer science from Beijing Jiaotong University, Beijing, China, in 2004, the M.Sc. degree in electronic engineering from the University of Surrey, Guildford, U.K., in 2006, and the Ph.D. degree in electronic engineering from the University of York, York, U.K., in 2010. From 2010 to 2011, he was a Post-Doctoral Fellow with the Electronics and Communications Laboratory, Conservatoire National des Arts et Metiers, Paris, France. Since 2011, he has been with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China, where he is currently an Associate Professor. His research interests include transceiver design for multiple-antenna systems, sensor array processing, adaptive filtering, full-duplex communications, co-operative and relay communications, and wireless information and energy transfer.



YULONG ZOU (SM'13) received the B.Eng. degree in information engineering from the Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China, in 2006, the first Ph.D. degree in electrical engineering from the Stevens Institute of Technology, NJ, USA, in 2012, and the second Ph.D. degree in signal and information processing from NUPT in 2012. He is currently a Professor with NUPT. His research interests span a wide range of topics in wireless communications and signal processing, including the co-operative communications, cognitive radio, wireless security, and energy-efficient communications.

Dr. Zou was a recipient of the 2014 IEEE Communications Society Asia-Pacific Best Young Researcher. He serves on the Editorial Board of the IEEE Communications Surveys and Tutorials, the IEEE COMMUNICATIONS LETTERS, the *I*ET Communications, and the *E*URASIP Journal on Advances in Signal Processing. In addition, he has served as symposium chairs, session chairs, and TPC members for a number of IEEE sponsored conferences, including the IEEE Wireless Communications and Networking Conference, the IEEE Global Communications Conference, the IEEE International Conference on Communications, the IEEE Vehicular Technology Conference, the International Conference on Communications in China.



LAJOS HANZO (M'91–SM'92–F'04) received the D.Sc. degree in electronics in 1976 and the Ph.D. degree in 1983. During his 40-year career in telecommunications he has held various research and academic posts in Hungary, Germany, and U.K. Since 1986, he has been with the School of Electronics and Computer Science, University of Southampton, U.K. He has successfully supervised 110 Ph.D. students. He is currently directing a 100-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council, U.K., the European IST Programme and the Mobile Virtual Centre of Excellence, U.K. His research is funded by the European Research Council's Senior Research Fellow Grant. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He has co-authored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published over 1600 research entries at the IEEE Xplore. He was FREng, FIET, and a fellow of the EURASIP. In 2009, he received the honorary doctorate Doctor Honoris Causa by the Technical University of Budapest and in 2015 by the University of Edinburgh. He served as the TPC Chair and the General Chair of the IEEE conferences, presented keynote lectures and has been received a number of distinctions. He is also a Governor of the IEEE VTS. From 2008 to 2012, he was the Editor-in-Chief of the IEEE Press and a Chaired Professor at Tsinghua University, Beijing. He is the Chair in telecommunications with the University of Southampton.

BENOIT CHAMPAGNE (S'87–M'89–SM'03) received the B.Ing. degree in engineering physics from the École Polytechnique de Montréal in 1983, the M.Sc. degree in physics from the Université de Montréal in 1985, and the Ph.D. degree in electrical engineering from the University of Toronto in 1990. From 1990 to 1999, he was an Assistant Professor and then an Associate Professor with INRS-Telecommunications, Université du Quebec, Montréal. In 1999, he joined McGill University, Montreal, where he is currently a Full Professor with the Department of Electrical and Computer Engineering. He served as an Associate Chairman of Graduate Studies with the Department from 2004 to 2007. His research focuses on the study of advanced algorithms for the processing of information bearing signals by digital means. His interests span many areas of statistical signal processing, including detection and estimation, sensor array processing, adaptive filtering, and applications thereof to broadband communications and audio processing, where he has co-authored over 200 referred publications. His research has been funded by the Natural Sciences and Engineering Research Council of Canada, the Fonds de Recherche sur la Nature et les Technologies from the Government of Quebec, and some major industrial sponsors, including Nortel Networks, Bell Canada, InterDigital, and Microsemi. He has been an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the *E*URASIP Journal on Applied Signal Processing. He has also served on the Technical Committees of several international conferences in the fields of communications and signal processing. In particular, he was the Co-Chair, Wide Area Cellular Communications Track, of the IEEE International Symposium on PIMRC, Toronto, ON, in 2011, the Co-Chair, Antenna and Propagation Track, of the IEEE VTC-Fall, Los Angeles, USA, in 2004, and the Registration Chair of the IEEE International Conference on ASSP, Montreal, QC, Canada, in 2004.