Understanding the Optimal Power Allocation Structure in AF-Based MIMO Cooperative Networks with Multiple Relays

Jia Liu[†] Ness B. Shroff[†] Hanif D. Sherali[‡]

[†]Department of Electrical and Computer Engineering, The Ohio State University

[‡]Grado Dept. of Industrial Systems Engineering, Virginia Polytechnic Institute and State University

Abstract

Cooperative networking is known to have significant potential in increasing network capacity and transmission reliability. Although there have been extensive studies on applying cooperative networking in multi-hop ad hoc networks, most works on cooperative network optimization are limited to the basic three-node relay scheme and single-antenna systems. These two limitations are interconnected and both due to a limited understanding of the optimal power allocation structure in MIMO cooperative networks (MIMO-CN). In this paper, we study in depth the structural properties of the optimal power allocation in MIMO-CN. More specifically, we show that the optimal power allocation at the source and each relay follows a *matching* structure in MIMO-CN. This result generalizes the power allocation result under the basic three-node setting to the multi-relay setting, for which the optimal power allocation structure has been heretofore unknown. We further quantify the performance gain due to cooperative relay and establish a connection between cooperative relay and pure relay. Finally, based on these structural insights, we reduce the MIMO-CN rate maximization problem to an equivalent scalar formulation. We then propose a global optimization method to solve this simplified and equivalent problem.

I. Introduction

A. Background and Motivation

The concept of cooperative networking [1], [2] can trace its roots back to the 1970s, when information-theoretic studies were first conducted in [3], [4] under the theme of "relay channels." In recent years, cooperative networking has received substantial interest from the wireless networking research community. Many interesting problems for cooperative networks have been actively researched, such as throughput-optimal scheduling [5], network lifetime maximization [6], distributed routing [7], MAC layer protocol design [8], just to name a few. Although there have been extensive studies concerning cooperative networks, most works on optimizing the performance of cooperative networks have the following major limitations:

i) Limited to the basic three-node relay scheme. The basic three-node relay scheme is shown in Fig. 1, where the message transmitted from the source S to the destination D is relayed by a node R, which can overhear

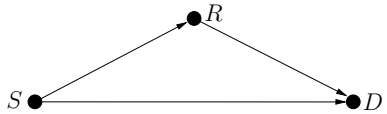


Fig. 1. The basic three-node relay scheme.

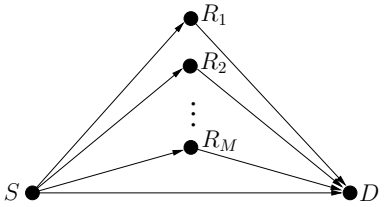


Fig. 2. A cooperative network with multiple relays.

the message. The benefit of such a cooperative transmission comes from the fact that the destination node can coherently combine two received signals coming from independent paths such that the transmission reliability can be greatly improved. In an ad hoc network environment, however, the message from the source is likely to be overheard by multiple neighboring nodes. A common cooperative communication approach in this situation is "relay assignment," i.e., we just choose one of the multiple neighboring nodes to be the relay node for which the three-node relay scheme can be applied (see, e.g., [9], [10] and references therein). Despite its simplicity, relay assignment is clearly suboptimal since all such neighboring nodes can potentially serve as relays to further improve the system performance, as shown in Fig. 2.

ii) Limited to single-antenna systems. In the current literature, research on cooperative networks with MIMO-enabled nodes remains very limited. In cooperative networks, it is very interesting to explore the idea of deploying multiple antennas at each node. This is because, with multiple antennas, the source and the relays can multiplex independent data streams by exploiting the inherent independent spatial channels. As a result, the end-to-end channel capacity can scale almost linearly as the number of antennas increases.

While the above two limitations are seemingly unrelated, they are in fact both associated with the limited understanding of the optimal power allocation structure of MIMO cooperative networks (MIMO-CN). To see why the first limitation is also related to MIMO, let us consider the multi-relay network in Fig. 2, where each node has a single antenna. In this example, we can treat all single-antenna relay nodes $R_1, \ldots, R_M$ as a single virtual relay node with M antennas. In that sense, analyzing this multi-relay network is closely related to analyzing a three-node cooperative network where the relay node is MIMO-enabled. Thus, we can see that besides the attractive capacity benefits of MIMO, studying MIMO-CN is of theoretical importance because it generalizes previous studies on single-antenna-based cooperative communication, which can be viewed as special cases of MIMO-CN.

In this paper, we explore the structural properties of optimal power allocation in MIMO-CN with multiple relays. More specifically, we consider the optimal power allocation structure at the source and each relay node so that the end-to-end achievable rate can be maximized.

B. Amplify-and-Forward (AF) or Decode-and-Forward (DF)?

Before studying MIMO-CN with multiple relays, an interesting and important question to answer is: *What cooperative relay strategy should we employ in MIMO-CN with multiple relays?* There are two main categories of cooperative relay strategies, namely, amplify-and-forward (AF) and decode-and-forward (DF) [1]. Simply speaking,

under AF, the relay node amplifies and retransmits the received signal without decoding the message. Under DF, the relay node decodes, re-encodes, and retransmits the signals. These two approaches are sometimes referred to as "non-regenerative" and "regenerative" strategies in the literature, respectively, and they both have pros and cons under different network settings.

In this work, our focus is on AF-based MIMO-CN. The first apparent reason is the lower complexity of AF since no decoding/encoding is needed under AF. This lower complexity is even more attractive in MIMO-CN, where decoding multiple data streams could be computationally intensive.

In addition to simplicity, a far more important reason for choosing AF in this work is because AF outperforms DF in the multi-relay setting. To see this, we again use the network in Fig. 2 as an example. We let SNR_{sd}, SNR_{sr_i} and $\mathsf{SNR}_{r_i d}$ represent the signal-to-noise ratios of the $S \to D$, $S \to R_i$ and the $R_i \to D$ links, respectively. Then, the end-to-end effective SNR under AF in this case is $\frac{M+1}{3}$, while the end-to-end effective SNR under DF is 1 (see Appendix A for details). Thus, it is clear that as long as $M > 2$, the achievable rate under AF is always higher than that under DF. In general, as the number of relays increases in MIMO-CN, the effective SNR under AF scales linearly, while the effective SNR under DF remains a constant.

C. Main Results and Contributions

The main results and contributions of this paper are as follows:

i) We show that the optimal power allocation at each relay follows a *matching* structure. More precisely, the diagonalization of each relay's power allocation matrix matches to certain eigen-directions of the joint source-relay and relay-destination channels. We further show that this matching structure is true regardless of the channel state information (CSI) assumption at the source, and regardless of the presence of the source-destination link. We point out that our result *generalizes* the matching structure under the basic three-node setting to the multi-relay setting, for which the optimal power allocation structure has been heretofore unknown. The proof of this result is not a trivial extension of that for the basic three-node relay setting because, as we shall see in Section III, the relay power allocation in the multi-relay setting is subject to a unique constraint.

ii) Through an analysis of the channel structures in MIMO-CN, we establish the relationship between MIMO-CN and multi-hop MIMO networks with pure relay links. More specifically, we quantify the performance gain due to cooperative relay. Not only does this result make the proof of the optimal power allocation structure in MIMO-CN easier, but it also builds a connection between cooperative relay and pure relay, advancing our understanding of the benefits of cooperative communications.

iii) Based on the structural insights of the optimal power allocation, we reduce the MIMO-CN rate maximization problem to an equivalent scalar formulation, which is similar to that of the conventional single-antenna three-node relay problem. We analyze the structure and convexity properties of the equivalent problem, and then propose a global optimization algorithm based on a branch-and-bound scheme coupled with a custom-designed convex-hull relaxation (BB/CHR), which *guarantees* finding a global optimal solution for this nonconvex problem.

D. Paper Organization

The remainder of this paper is organized as follows. Section II discusses the related work, putting our work in a comparative perspective. Section III presents the network model and problem formulation. In Section IV and Section V, we study the optimal power allocation structures corresponding to pure relay and cooperative relay paradigms, and point out their connections. Based on these structural results, we simplify and reformulate the underlying optimization problems. In Section VI, we propose a novel global optimization method based on the BB/CHR framework to solve the reformulated equivalent problem. Numerical results are also provided to show the efficacy of the proposed algorithms. Section VII concludes this paper.

II. Related Work

Since the benefits of cooperative networking were recognized [1], [2], several initial attempts on extending cooperative networking to MIMO have been reported in the literature [11]–[16]. In [11], Tang and Hua first considered the optimal relay amplifying matrix for the basic three-node MIMO-CN under the assumption that the source-relay CSI is unknown. Their main conclusion is that when the direct link between the source and the destination is not present (i.e., pure relay), the optimal amplifying matrix adopts a "matching" structure. Coincidentally, Muñoz-Medina *et al.* [12] independently arrived at the same conclusion via a different proof technique. Later in [17], Fang *et al.* generalized the matching result to the three-node MIMO-CN network where the source has full CSI. Recent works on MIMO-CN started to consider more complex relay settings. In [18], Fu *et al.* studied MIMO-CN with multiple AF relays, which is similar to our setting. However, their work differs from ours in that they assumed a sum power constraint across all relay nodes. This assumption is usually not realistic since each relay has its own power budget. Thus, a power constraint on each individual node is more appropriate. As we shall see later, imposing an individual power constraint on each relay node results in a more challenging power allocation problem. It is worth pointing out that the three-node multi-carrier MIMO-CN considered in [15] can also be viewed as a MIMO-CN with multiple relays. Compared to our network setting, the major difference is that each source-relay-destination path in [19] operates using orthogonal channels (subcarriers) that do not cooperate with each other. This setting yields a more tractable problem, which can be thought of as a special case of the model we consider in this paper.

III. Network Model and Problem Formulation

We first introduce the notation used in this paper. We use boldface to denote matrices and vectors. For a complex-valued matrix $\mathbf{M}$, we let $\mathbf{M}^*$ and $\mathbf{M}^\dagger$ denote the conjugate and conjugate transpose of $\mathbf{M}$, respectively. $\mathrm{Tr}\{\mathbf{M}\}$ denotes the trace of $\mathbf{M}$. We let $\mathbf{I}_N$ denote the $N \times N$ identity matrix. $\mathbf{M} \succeq 0$ represents that $\mathbf{M}$ is Hermitian and positive semidefinite (PSD). $\mathrm{Diag}\{\mathbf{M}_1 \dots \mathbf{M}_n\}$ represents the block diagonal matrix with matrices $\mathbf{M}_1, \dots, \mathbf{M}_n$ on its main diagonal. $\mathbf{M} \circ \mathbf{N}$ represents the Hadamard product of matrices $\mathbf{M}$ and $\mathbf{N}$. We let $(\mathbf{M})_{ij}$ denote the entry in the i-th row and j-th column in matrix $\mathbf{M}$ and let $(\mathbf{v})_j$ denote the j-th entry in vector $\mathbf{v}$.

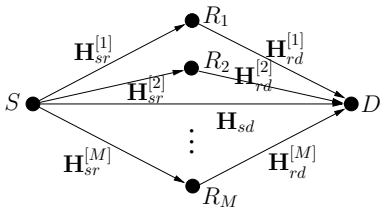


Fig. 3. An AF-based MIMO-CN with M relay nodes.

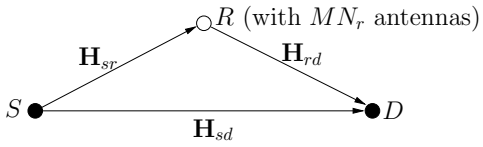


Fig. 4. An AF-based MIMO-CN with M relay nodes grouped as one virtual super relay node.

We consider an AF-based MIMO-CN as shown in Fig. 3. The source node S transmits messages to the destination node D, assisted by M relay nodes R_1, R_2, ..., R_M. The source and the destination nodes have N_s and N_d antennas, respectively. For ease of exposition, we assume that each relay node has N_r antennas in this paper. We note that the generalization to cases where each relay node has distinct number of antennas can also be incorporated in our subsequent analysis in a similar fashion but at the expense of more complicated notation. We let $\mathbf{H}_{sr}^{[i]} \in \mathbb{C}^{N_r \times N_s}$ denote the channel gain matrix between the source and the i-th relay. We let $\mathbf{H}_{rd}^{[i]} \in \mathbb{C}^{N_d \times N_r}$ denote the channel gain matrix between the i-th relay and the destination. The channel gain matrix for the direct link between the source and the destination is denoted by $\mathbf{H}_{sd} \in \mathbb{C}^{N_d \times N_s}$.

Due to self-interference, each relay node cannot transmit and receive in the same channel at the same time (i.e., we operate in a half-duplex mode). Thus, a transmission in an AF-based MIMO-CN takes two time slots. In the first time slot, the source transmits a message to the destination node. Due to the broadcast nature of wireless media, the same message can be overheard by all relays. In the second time slot, each relay simply amplifies and transmits its received signal to the destination without decoding the message. At the end of the second time slot, the destination coherently combines all received signals to decode the message.

Throughout the rest of the paper, we classify MIMO-CN into two cases depending on whether or not the direct channel between S and D is strong enough to support communication. We refer to the case where the direct channel is absent as "pure relay" (PR) and refer to the normal case as "cooperative relay" (CR). Although PR is a special case of CR, this categorization proves to be useful in that: 1) it is easier to obtain structural properties under the simpler PR case and these properties provide important insights to the more complex CR case; 2) this categorization helps build a connection between PR and CR and deepens our understanding of cooperation benefits.

A. Input-Output Signal Model

1) Source-Relay: In a MIMO-CN, the received signal at the i-th relay $\mathbf{y}_r^{[i]}$ can be written as

$$\mathbf{y}_r^{[i]} = \mathbf{H}_{sr}^{[i]}\mathbf{x}_s + \mathbf{n}_r^{[i]}, \quad i = 1, 2, \ldots, M, \tag{1}$$

where $\mathbf{x}_s \in \mathbb{C}^{N_s \times 1}$ represents the transmit signal vector and $\mathbf{n}_r^{[i]} \in \mathbb{C}^{N_r \times 1}$ represents the zero-mean circularly symmetric Gaussian noise vector seen at the i-th relay.

For notational convenience, we can treat the relay nodes $R_1, \ldots, R_M$ as a virtual super relay node R with MN_r antennas, as shown in Fig. 4. This equivalent form would allow us to formulate an AF-based MIMO-CN in a more

compact form. However, we emphasize that although this equivalent form appears to be similar to the three-node relay networks, the unique constraint on the amplification factor yields a different and more challenging problem (more on this later). As shown in Fig. 4, we let $\mathbf{H}_{sr} \triangleq \left[\mathbf{H}_{sr}^{[1]\dagger}, \ldots, \mathbf{H}_{sr}^{[M]\dagger}\right]^{\dagger} \in \mathbb{C}^{MN_r \times N_s}$ represent the combined relay channel matrix between the source and the relays. Also, we let $\mathbf{n}_r \triangleq [\mathbf{n}_r^{[1]\dagger}, \ldots, \mathbf{n}_r^{[M]\dagger}]^{\dagger} \in \mathbb{C}^{MN_r \times 1}$. Then, the combined received signal at the virtual super relay node, defined as $\mathbf{y}_r \triangleq [\mathbf{y}_r^{[1]\dagger}, \ldots, \mathbf{y}_r^{[M]\dagger}]^{\dagger} \in \mathbb{C}^{MN_r}$, can be conveniently written as $\mathbf{y}_r = \mathbf{H}_{sr}\mathbf{x}_s + \mathbf{n}_r$.

2) Amplification-and-Forward: Recall that in an AF-based MIMO-CN, each relay amplifies the received signal. Compared to conventional single-antenna AF-based cooperative networks (CN), a major difference in AF-based MIMO-CN is the form of the *amplification factor*. Unlike single-antenna CN where the amplification factor can be represented by a scalar, in multi-relay MIMO-CN, the general form of the amplification factor at the i-th relay should be represented by a *matrix* $\mathbf{A}_i \in \mathbb{C}^{N_r \times N_r}$. That is, the relay signal at the i-th relay node $\mathbf{x}_r^{[i]}$ can be written as $\mathbf{x}_r^{[i]} = \mathbf{A}_i\mathbf{y}_r^{[i]}$. Physically, this means that the transmit signal at each antenna in the second time slot is a *sum of amplified received signals from all antennas* in the first time slot. To see this, we can expand each entry in the transmit signal $\mathbf{x}_r^{[i]} = \mathbf{A}_i\mathbf{y}_r^{[i]}$ as

$$(\mathbf{x}_r^{[i]})_j = \sum_{k=1}^{N_r} (\mathbf{A}_i)_{jk}(\mathbf{y}_r^{[i]})_k, \quad j = 1, 2, \ldots, N_r. \tag{2}$$

It is worth pointing out that the AF strategy is completely determined by the *structure* of $\mathbf{A}_i$. For example, if $\mathbf{A}_i$ is a diagonal matrix, then it can be seen from (2) that each antenna simply amplifies and retransmits its own received signal (i.e., there is no summation of signals from other antennas). Since the structure of each $\mathbf{A}_i$ could significantly impact the performance of an AF-based MIMO-CN, one of the main goals in this paper is to understand the optimal structural property of the amplification matrix at each relay node.

Due to the *distributed nature* of ad hoc network environments, the relay nodes cannot share their received signals with each other. As a result, although we can represent a multi-relay MIMO-CN in an equivalent form as shown in Fig. 4, the virtual node R does not really function as a single node. In other words, each R_i within R cannot amplify and forward other relay nodes' signals because they do not share antennas. Mathematically, this can be modeled by introducing a *block diagonal constraint* on $\mathbf{A}$, which denotes the amplification matrix for the virtual super relay node R in Fig. 4, i.e.,

$$\mathbf{A} = \text{Diag}\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_M\} \in \mathbb{C}^{MN_r \times MN_r}. \tag{3}$$

With matrix $\mathbf{A}$, we can represent the transmit signal of the virtual super relay node as $\mathbf{x}_r = \mathbf{A}\mathbf{y}_r = \mathbf{A}(\mathbf{H}_{sr}\mathbf{x}_s + \mathbf{n}_r)$.

3) Relay-Destination: As shown in Fig. 4, we use a matrix $\mathbf{H}_{rd} \triangleq \left[\mathbf{H}_{rd}^{[1]}, \ldots, \mathbf{H}_{rd}^{[M]}\right] \in \mathbb{C}^{N_d \times MN_r}$ to represent the combined channel gain matrix between the relays and the destination. In the PR case, the received signal at the destination node $\mathbf{y}_d$ can be written as

$$\mathbf{y}_d = \mathbf{H}_{rd}\mathbf{x}_r + \mathbf{n}_d^{(2)} = \mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr}\mathbf{x}_s + (\mathbf{H}_{rd}\mathbf{A}\mathbf{n}_r + \mathbf{n}_d^{(2)}), \tag{4}$$

where $\mathbf{n}_d^{(2)} \in \mathbb{C}^{N_d}$ is a zero-mean circularly symmetric Gaussian noise vector seen at D in the second time slot.

In the CR case, since a direct channel is present, we can expand the received signal $\mathbf{y}_d$ in (4) as follows:

$$\mathbf{y}_d = \begin{bmatrix} \mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix} \mathbf{x}_s + \begin{bmatrix} \mathbf{H}_{rd}\mathbf{A} & \mathbf{I}_{N_d} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_d} \end{bmatrix} \begin{bmatrix} \mathbf{n}_r \\ \mathbf{n}_d^{(2)} \\ \mathbf{n}_d^{(1)} \end{bmatrix}, \tag{5}$$

where $\mathbf{n}_d^{(1)} \in \mathbb{C}^{N_d}$ denotes the zero-mean circularly symmetric Gaussian noise vector seen at D in the first time slot.

B. Problem Formulation

1) Achievable Rate Computation: Under PR, the achievable rate of an AF-based MIMO-CN can be computed as

$$I_{PR}(\mathbf{Q}, \mathbf{A}) \;=\; \frac{1}{2}\log_2 \left| \mathbf{I}_{N_d} + (\mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr})\mathbf{Q}(\mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr})^\dagger (\sigma_d^2 \mathbf{I}_{N_d} + \sigma_r^2 \mathbf{H}_{rd}\mathbf{A}\mathbf{A}^\dagger \mathbf{H}_{rd}^\dagger)^{-1} \right|, \tag{6}$$

where $\mathbf{Q} \triangleq \mathbb{E}\{\mathbf{x}\mathbf{x}^\dagger\}$ represents the input signal covariance matrix (i.e., the source power); and σ_r^2 and σ_d^2 denote the variances of $\mathbf{n}_r$ and $\mathbf{n}_{rd}^{(2)}$, respectively. The factor $\frac{1}{2}$ in (6) accounts for the fact that two time slots are required to complete a transmission. By letting $\bar{\mathbf{H}}_{sd} \triangleq \mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr}$ denote the equivalent end-to-end channel under PR and letting $\bar{\mathbf{R}} \triangleq \sigma_d^2 \mathbf{I}_{N_d} + \sigma_r^2 \mathbf{H}_{rd}\mathbf{A}\mathbf{A}^\dagger \mathbf{H}_{rd}^\dagger$ denote the equivalent noise power at the destination, we can compactly rewrite (6) as

$$I_{PR}(\mathbf{Q}, \mathbf{A}) = \frac{1}{2}\log_2 \left| \mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^\dagger \bar{\mathbf{R}}^{-1} \right|. \tag{7}$$

Similarly, we can write the achievable rate under CR as follows:

$$I_{CR}(\mathbf{Q}, \mathbf{A}) = \frac{1}{2}\log_2 \left| \mathbf{I}_{2N_d} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger \mathbf{R}^{-1} \right|. \tag{8}$$

In (8), the equivalent end-to-end channel gain matrix $\mathbf{H} \in \mathbb{C}^{2N_d \times N_s}$ under CR is defined as

$$\mathbf{H} \triangleq \begin{bmatrix} \bar{\mathbf{H}}_{sd} \\ \mathbf{H}_{sd} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{rd}\mathbf{A}\mathbf{H}_{sr} \\ \mathbf{H}_{sd} \end{bmatrix},$$

and the equivalent noise power $\mathbf{R} \in \mathbb{C}^{2N_d \times 2N_d}$ is defined as

$$\mathbf{R} \triangleq \begin{bmatrix} \bar{\mathbf{R}} & \mathbf{0} \\ \mathbf{0} & \sigma_d^2 \mathbf{I}_{N_d} \end{bmatrix} = \begin{bmatrix} \sigma_d^2 \mathbf{I}_{N_d} + \sigma_r^2 \mathbf{H}_{rd}\mathbf{A}\mathbf{A}^\dagger \mathbf{H}_{rd}^\dagger & \mathbf{0} \\ \mathbf{0} & \sigma_d^2 \mathbf{I}_{N_d} \end{bmatrix}.$$

2) Power Constraints: Due to the maximum transmit power limit at the source and at each relay node, we have the following power constraints for the source and each relay node: $\mathrm{Tr}(\mathbf{Q}) \le P_T$, $\mathrm{Tr}(\mathbf{A}_i(\sigma_r^2 \mathbf{I}_{N_r} + \mathbf{H}_{sr}\mathbf{Q}\mathbf{H}_{sr}^\dagger)\mathbf{A}_i^\dagger) \le P_R$, $i = 1, \ldots, K$, where P_T and P_R represent the maximum transmit power limit at the source node and at each relay node, respectively. For compactness, we define the following constraint sets:

$$\Omega \triangleq \{\mathbf{Q} \,|\mathrm{Tr}(\mathbf{Q}) \le P_T\},$$
$$\Psi \triangleq \left\{ \mathrm{Diag}\{\mathbf{A}_1, \ldots, \mathbf{A}_M\} \;\middle|\; \begin{array}{l} \mathrm{Tr}(\mathbf{A}_i(\sigma_r^2 \mathbf{I}_{N_r} + \mathbf{H}_{sr}\mathbf{Q}\mathbf{H}_{sr}^\dagger) \\ \mathbf{A}_i^\dagger) \le P_R, \forall i = 1, 2, \ldots, K \end{array} \right\}.$$

3) Problem Formulation: To maximize the end-to-end achievable rate, we need to find an optimal $\mathbf{Q}$ and $\mathbf{A}$ to maximize $I_{PR}(\mathbf{Q},\mathbf{A})$ under PR or $I_{CR}(\mathbf{Q},\mathbf{A})$ under CR. This can be formulated as the following joint source-relay power optimization (PO) problem under PR and CR, respectively:

$$\begin{aligned} \underline{\textbf{PO-PR}}\textbf{:}\ \text{Maximize} \quad & I_{PR}(\mathbf{Q},\mathbf{A}) \\ \text{subject to} \quad & \mathbf{Q}\in\Omega, \mathbf{A}\in\Psi, \end{aligned} \tag{9}$$

$$\begin{aligned} \underline{\textbf{PO-CR}}\textbf{:}\ \text{Maximize} \quad & I_{CR}(\mathbf{Q},\mathbf{A}) \\ \text{subject to} \quad & \mathbf{Q}\in\Omega, \mathbf{A}\in\Psi. \end{aligned} \tag{10}$$

We remark that both PO-PR and PO-CR can be naturally decomposed into two parts as follows: $\max_{\mathbf{Q}\in\Omega} I_j^*(\mathbf{Q})$, where $I_j^*(\mathbf{Q}) \triangleq \max_{\mathbf{A}\in\Psi} I_j(\mathbf{Q},\mathbf{A})$, and where $j\in\{PR,CR\}$. Hence, solving PO-PR and PO-CR boils down to solving an inner subproblem with respect to $\mathbf{A}$ when $\mathbf{Q}$ is held fixed and an outer main program with respect to $\mathbf{Q}$. Due to the complex matrix variables and operations, directly tackling such problems is intractable in general. For PO-PR and PO-CR, however, it turns out that we can exploit the inherent special structure to significantly reduce the complexity of the problems. Thus, in what follows, we will first study the structural properties of the optimal $\mathbf{A}$ and $\mathbf{Q}$ in PO-PR and PO-CR. Based on these properties, we will reformulate PO-PR and PO-CR. We will start with the relatively simpler PO-PR problem. It will soon be clear that the results under the PR case provide fundamental insights into the more complex PO-CR problem.

IV. Optimal Power Allocation Structure: The Pure Relay Case

In this section, we will first investigate the structural properties of an optimal $\mathbf{A}$ in Section IV-A. Then, we will study the structural properties of an optimal $\mathbf{Q}$ in Section IV-B. Based on these structural properties, we will reformulate the PO-PR problem in Section IV-C.

A. The Structure of An Optimal Amplification Matrix

For now, we assume that $\mathbf{Q}$ is known. To expose the structure of $I_{PR}(\mathbf{Q},\mathbf{A})$, we let $\tilde{\mathbf{H}}_{rd} \triangleq \frac{\sigma_r}{\sigma_d}\mathbf{H}_{rd}\mathbf{A}$ and $\tilde{\mathbf{H}}_{sr} \triangleq \frac{1}{\sigma_r}\mathbf{H}_{sr}\mathbf{Q}^{\frac{1}{2}}$. To study the structural properties of an optimal $\mathbf{A}$, we first need the following result to simplify $I_{PR}(\mathbf{Q},\mathbf{A})$.

Lemma 1. *The achievable rate expression in (6) is equivalent to the following expression:*

$$I_{PR}(\mathbf{Q},\mathbf{A}) = \frac{1}{2}\log_2\left|\mathbf{I}_{MN_r} + \tilde{\mathbf{\Lambda}}_{sr}(\mathbf{I}_{MN_r} - (\mathbf{I}_{MN_r} + \tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr})^{-1})\right|, \tag{11}$$

where the real diagonal matrix $\tilde{\mathbf{\Lambda}}_{sr}\in\mathbb{C}^{MN_r\times MN_r}$ and the orthonormal matrix $\tilde{\mathbf{U}}_{sr}\in\mathbb{C}^{MN_r\times MN_r}$ are obtained from the eigenvalue decomposition of $\tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger}$, i.e., $\tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger} = \tilde{\mathbf{U}}_{sr}\tilde{\mathbf{\Lambda}}_{sr}\tilde{\mathbf{U}}_{sr}^{\dagger}$.

Proof: See Appendix B. ■

Lemma 1 implies that maximizing (6) with respect to $\mathbf{A}$ can be equivalently done on (11), which is relatively easier due to the *diagonal* structure in (11). More specifically, according to the Hadamard inequality [20], the

achievable rate in (11) is maximized when the matrix inside the determinant is diagonal. Note that in (11), every matrix term is already diagonal except the term $(\mathbf{I}_{MN_r} + \tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr})^{-1}$. Thus, for the overall matrix to be diagonal, it suffices that $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr}$ is diagonal. To this end, we have the following sufficient condition for the diagonality of $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr}$.

Lemma 2. *$\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr}$ is diagonal if $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$ is diagonal, where $\mathbf{V}_{rd}$ is the eigenvector matrix for $\mathbf{H}_{rd}^{\dagger}\mathbf{H}_{rd}$ i.e., $\mathbf{H}_{rd}^{\dagger}\mathbf{H}_{rd} = \mathbf{V}_{rd}\mathbf{\Lambda}_{rd}\mathbf{V}_{rd}^{\dagger}$.*

Proof: To see this, we rewrite $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr}$ in the following form: $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr} = (\sigma_r^2/\sigma_d^2)\tilde{\mathbf{U}}_{sr}^{\dagger}\mathbf{A}^{\dagger}\mathbf{H}_{rd}^{\dagger}\mathbf{H}_{rd}\mathbf{A}\tilde{\mathbf{U}}_{sr} = (\sigma_r^2/\sigma_d^2)\tilde{\mathbf{U}}_{sr}^{\dagger}\mathbf{A}^{\dagger}\mathbf{V}_{rd}\mathbf{\Lambda}_{rd}\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$. Since $\mathbf{\Lambda}_{rd}$ is diagonal, in order for $\tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr}$ to be diagonal, it suffices for $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$ to be diagonal. ■

Now, recall that $\mathbf{A} = \mathrm{Diag}\{\mathbf{A}_1, \ldots, \mathbf{A}_M\}$. Thus, $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$ can be expanded as

$$\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr} = \sum_{k=1}^{M}(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]}, \tag{12}$$

where $\mathbf{V}_{rd}^{[k]} \in \mathbb{C}^{N_r \times MN_r}$ and $\tilde{\mathbf{U}}_{sr}^{[k]} \in \mathbb{C}^{N_r \times MN_r}$ represent the submatrices in $\mathbf{V}_{rd}$ and $\tilde{\mathbf{U}}_{sr}$ starting from the $((k-1)N_r+1)$-st row to the kN_r-th row, respectively, i.e.,

$$\mathbf{V}_{rd} = \left[\begin{array}{c} \mathbf{V}_{rd}^{[1]} \\ \hline \vdots \\ \hline \mathbf{V}_{rd}^{[M]} \end{array}\right] \quad \text{and} \quad \tilde{\mathbf{U}}_{sr} = \left[\begin{array}{c} \tilde{\mathbf{U}}_{sr}^{[1]} \\ \hline \vdots \\ \hline \tilde{\mathbf{U}}_{sr}^{[M]} \end{array}\right].$$

Based on (12), we have the first main result in this paper.

Theorem 1. *When $\mathbf{Q}$ is fixed, the relay amplification matrix $\mathbf{A}_k$ at the k-th relay node has the following structure at optimality:*

$$\mathbf{A}_k = ((\mathbf{V}_{rd}^{[k]})^{\dagger})^{\mathrm{LI}}\mathbf{\Phi}_k(\tilde{\mathbf{U}}_{sr}^{[k]})^{\mathrm{RI}}, \quad k = 1, 2, \ldots, M, \tag{13}$$

where $(\cdot)^{\mathrm{LI}}$ and $(\cdot)^{\mathrm{RI}}$ represent the left inverse and the right inverse, respectively; and $\mathbf{\Phi}_k \in \mathbb{R}^{MN_r \times MN_r}$ is real diagonal.

Theorem 1 is important because, as we shall see later, the diagonal decomposition structure in (13) significantly simplifies the determinant calculation and allows us to reduce the matrix-based problem into a scalar form. The proof of Theorem 1 relies on the following key lemma.

Lemma 3. *$\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$ is diagonal if and only if for all $k = 1, \ldots, M$, $(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]}$ is diagonal.*

Proof: The "if" part follows immediately from (12). For the "only if" part, Lemma 3 can be proved by first assuming that not all $(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]}$ are diagonal, $k = 1, \ldots, M$. Then, by analyzing a homogeneous linear equation system associated with $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$, we can reach a contradiction.

First, notice that since $\mathbf{V}_{rd}$ is orthonormal, any two row vectors selected from $\mathbf{V}_{rd}^{[i]}$ and $\mathbf{V}_{rd}^{[j]}$, $i \neq j$, are linearly independent. Next, suppose that each $(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]}$ is possibly non-diagonal. By expanding (12) entry-wise and

noting that the non-diagonal entry of $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$ is zero (since $\mathbf{V}_{rd}^{\dagger}\mathbf{A}\tilde{\mathbf{U}}_{sr}$ is diagonal), we have

$$\sum_{i=1}^{N_r}\sum_{j=1}^{N_r} v_{pi}^{[1]} a_{ij}^{[1]} \tilde{u}_{jq}^{[1]} + \cdots + \sum_{i=1}^{N_r}\sum_{j=1}^{N_r} v_{pi}^{[M]} a_{ij}^{[M]} \tilde{u}_{jq}^{[M]} = 0, \tag{14}$$

for all $p, q = 1, 2, \ldots, MN_r$ and $p \neq q$. In (14), $v_{pi}^{[i]}$ represents the entry in the p-th row and i-th column of the matrix $\mathbf{V}_{rd}^{[k]}$, $a_{ij}^{[k]}$ represents the entry in the i-th row and j-th column of the matrix $\mathbf{A}_i^{[k]}$, and $\tilde{u}_{jq}^{[k]}$ represents the entry in the j-th row and q-th column of the matrix $\tilde{\mathbf{U}}_{sr}^{[k]}$. Since there are a total of $M^2N_r^2 - MN_r$ such linearly independent equations in the form of (14) with MN_r^2 unknowns (i.e., $a_{ij}^{[k]}$-variables); in order for this homogeneous linear equation system to have a non-zero solution, the dimension of the null space needs to be greater than zero, i.e., $M^2N_r^2 - MN_r < MN_r^2$. This implies that

$$N_r < \frac{1}{M-1}. \tag{15}$$

However, note that in a multi-relay setting we have, $M \geq 2$. Also, N_r (i.e., the number of antennas at the relays) has to be an integer, which implies that the inequality in (15) cannot be true – a contradiction. ■

Using Lemma 3, we are now ready to prove Theorem 1.

Proof of Theorem 1: Since each $(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]}$ term is a diagonal matrix, we let

$$(\mathbf{V}_{rd}^{[k]})^{\dagger}\mathbf{A}_k\tilde{\mathbf{U}}_{sr}^{[k]} = \mathbf{\Phi}_k, \tag{16}$$

where $\mathbf{\Phi}_k$, as stated in the theorem, is a diagonal matrix. Since $(\mathbf{V}_{rd}^{[k]})^{\dagger}$ and $\tilde{\mathbf{U}}_{sr}^{[k]}$ are tall-skinny and short-fat, there exist left and right inverses for $(\mathbf{V}_{rd}^{[k]})^{\dagger}$ and $\tilde{\mathbf{U}}_{sr}^{[k]}$, respectively. Then, Eq.(13) simply follows from multiplying $((\mathbf{V}_{rd}^{[k]})^{\dagger})^{\mathrm{LI}}$ and $(\tilde{\mathbf{U}}_{sr}^{[k]})^{\mathrm{RI}}$ on the left- and right- hand sides of (16), respectively. ■

Remark 1. We can see from (13) that $\mathbf{A}_k$ contains two parts: $(\tilde{\mathbf{U}}_{sr}^{[k]})^{\mathrm{RI}}$ matching to the eigen-directions of the joint source-relay channel, and $(\mathbf{V}_{rd}^{[k]})^{\mathrm{LI}}$ matching to the eigen-directions of the joint relay-destination channel. When the direction matching is done, the actual power allocation is completely determined by the diagonal entries in $\mathbf{\Phi}_k$. We point out that by imposing the individual power constraint, the structure of $\mathbf{A}_k$ is drastically different from that in [18], where all relays are subject to a sum power constraint.

Remark 2. It is worth pointing out that Theorem 1 implies that each relay needs the knowledge of the joint channel $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$. This means that, although the relays do not share received signals, they do need to share CSI, which is the price to pay in order to achieve optimal performance. This CSI requirement could potentially impose some challenges on practical implementations, and limited CSI feedback could be desirable in practice.

B. The Structure of Optimal Source Covariance Matrix

From (11), it can be seen that after choosing an appropriate $\mathbf{A}$, the value of the achievable rate depends on the diagonal entry in $\tilde{\mathbf{\Lambda}}_{sr}$. Thus, following a similar argument as in [17], [21], it can be shown that an optimal $\mathbf{Q}$ should match with the dominant right singular matrix of $\mathbf{H}_{sr}$. We state the result in the following proposition and omit its proof in this paper.

Proposition 1. *The matrix* $\mathbf{Q}$ *has the following structure at optimality:* $\mathbf{Q} = \mathbf{V}_{sr}\mathbf{\Lambda}\mathbf{V}_{sr}^{\dagger}$*, where* $\mathbf{V}_{sr}$ *represents the right singular matrix of the combined channel matrix* $\mathbf{H}_{sr}$*, i.e.,* $\mathbf{H}_{sr} = \mathbf{U}_{sr}\mathbf{\Sigma}_{sr}\mathbf{V}_{sr}^{\dagger}$*, and where* $\mathbf{\Sigma}_{sr}$ *is a diagonal matrix, and* $\mathbf{U}_{sr}$ *and* $\mathbf{V}_{sr}$ *are the associated left and right singular matrices, respectively.*

C. Problem Reformulation

Since the optimal power allocation for MIMO-CN is determined by the diagonal entries after appropriate matchings for $\mathbf{A}$ and $\mathbf{Q}$ are accomplished, the optimal power allocation problem can be simplified to a scalar form. Here, we first simplify the objective function in (11). After some algebraic manipulation from Theorem 1, we can obtain that

$$\begin{aligned} I_{PR}(\mathbf{Q},\mathbf{A}) &= \frac{1}{2}\log_2\Big|\mathbf{I}_{MN_r} + \tilde{\mathbf{\Lambda}}_{sr}(\mathbf{I}_{MN_r} - (\mathbf{I}_{MN_r} + \\ &\quad \frac{\sigma_r^2}{\sigma_d^2}\mathrm{Diag}\Big\{\lambda_{rd}^{(1)}(\textstyle\sum_{i=1}^{M} d_1^{(i)})^2,\ldots,\lambda_{rd}^{(MN_r)}(\textstyle\sum_{i=1}^{M} d_{MN_r}^{(i)})^2\Big\})^{-1})\Big| \qquad (17) \\ &= \frac{1}{2}\sum_{j=1}^{MN_r}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}(\sum_{i=1}^{M} d_j^{(i)})^2}{\lambda_{rd}^{(j)}(\sum_{i=1}^{M} d_j^{(i)})^2 + \frac{\sigma_d^2}{\sigma_r^2}}\right]. \qquad (18) \end{aligned}$$

Here, the eigenvalues $\{\tilde{\lambda}_{sr}^{(j)}\}_{j=1}^{MN_r}$ and $\{\lambda_{rd}^{(j)}\}_{j=1}^{MN_r}$ are both sorted in non-increasing order. Also, the source power constraint can be re-written as

$$\sum_{i=1}^{N_s}\tilde{\lambda}_{sr}^{(i)} \le P_T. \qquad (19)$$

By using the matrix identity $\mathrm{Tr}(\mathbf{MXNX}^{\dagger}) = \mathbf{x}^{\dagger}(\mathbf{M}\circ\mathbf{N}^{T})\mathbf{x}$, where $\mathbf{M}$ and $\mathbf{N}$ are square matrices; and $\mathbf{X}$ is a diagonal matrix with $\mathbf{x}$ on its main diagonal [18], [22], it follows that

$$\mathrm{Tr}(\mathbf{A}_i(\sigma_r^2\mathbf{I}_{N_r} + \mathbf{H}_{sr}\mathbf{Q}\mathbf{H}_{sr}^{\dagger})\mathbf{A}_i^{\dagger}) \le P_R \Rightarrow \mathbf{d}_i^{\dagger}\mathbf{S}_i\mathbf{d}_i \le P_R, \qquad (20)$$

where $\mathbf{S}_i$ is defined as

$$\mathbf{S}_i = \left[(((\mathbf{V}_{rd}^{[i]})^{\dagger})^{\mathrm{LI}})^{\dagger}((\mathbf{V}_{rd}^{[i]})^{\dagger})^{\mathrm{LI}}\right]\circ\left[(\mathbf{U}_{sr}^{[i]})^{\mathrm{RI}}(\sigma_r^2\mathbf{I}_{N_r} + \mathbf{H}_{sr}\mathbf{Q}\mathbf{H}_{sr}^{\dagger})((\mathbf{U}_{sr}^{[i]})^{\mathrm{RI}})^{\dagger}\right]^{T} \in \mathbb{C}^{MN_r\times MN_r}.$$

Note that $\mathbf{S}_i$ is symmetric and positive semidefinite because it is a Hadamard product of two positive semidefinite matrices.

Putting together all the above transformations and letting $\delta_j \triangleq \sum_{i=1}^{M} d_j^{(i)}$, we can equivalently write the PO-PR

problem in the following form:

PO-PR-SCALAR:

$$\begin{aligned}
\text{Maximize } & \frac{1}{2}\sum_{j=1}^{MN_r}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}\delta_j^2}{\lambda_{rd}^{(j)}\delta_j^2+\frac{\sigma_d^2}{\sigma_r^2}}\right] \\
\text{subject to } & \sum_{j=1}^{MN_r}(\mathbf{S}_i)_{jj}(d_j^{(i)})^2+\sum_{j=1}^{MN_r}\sum_{k=1,k\neq j}^{MN_r}(\mathbf{S}_i)_{jk}d_j^{(i)}d_k^{(i)}\leq P_R,\ i=1,\ldots,M, \\
& \delta_j-\sum_{i=1}^{M}d_j^{(i)}=0,\quad j=1,\ldots,MN_r, \\
& \sum_{i=1}^{N_s}\tilde{\lambda}_{sr}^{(i)}\leq P_T,
\end{aligned}$$

where the decision variables are δ_j, $d_j^{(i)}$, and $\tilde{\lambda}_{sr}^{(i)}$, $\forall i,j$.

Moreover, observe that $\tilde{\lambda}_{sr}\lambda_{rd}^{(j)}=0$ for $j=D+1,\ldots,MN_r$, where $D\triangleq\min\{rank(\tilde{\mathbf{H}}_{sr}),rank(\mathbf{H}_{rd})\}$. Therefore, the objective function can be rewritten as $\frac{1}{2}\sum_{j=1}^{D}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}\delta_j^2}{\lambda_{rd}^{(j)}\delta_j^2+\frac{\sigma_d^2}{\sigma_r^2}}\right]$. From the discussions regarding the objective function, it follows that PO-PR-SCALAR can be further simplified into an equivalent problem as given below:

PO-PR-SIM:

$$\begin{aligned}
\text{Maximize } & \frac{1}{2}\sum_{j=1}^{D}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}\delta_j^2}{\lambda_{rd}^{(j)}\delta_j^2+\frac{\sigma_d^2}{\sigma_r^2}}\right] \\
\text{subject to } & \sum_{j=1}^{MN_r}(\mathbf{S}_i)_{jj}(d_j^{(i)})^2+\sum_{j=1}^{MN_r}\sum_{k=1,k\neq j}^{MN_r}(\mathbf{S}_i)_{jk}d_j^{(i)}d_k^{(i)}\leq P_R,\ i=1,\ldots,M, \\
& \delta_j-\sum_{i=1}^{M}d_j^{(i)}=0,\quad j=1,\ldots,MN_r, \\
& \sum_{i=1}^{N_s}\tilde{\lambda}_{sr}^{(i)}\leq P_T,
\end{aligned}$$

where the decision variables are δ_j, $d_j^{(i)}$, and $\tilde{\lambda}_{sr}^{(i)}$, $\forall i,j$.

Remark 3. We note that matrix variables no longer appear in PO-PR-SIM, which significantly simplifies the formulation and reduces the computational complexity. Note also that PO-PR-SIM is in a mathematical form similar to that for a single-antenna CN. This suggests that solutions to PO-PR-SIM may be developed by drawing upon the rich experiences gained for single-antenna CN. Indeed, we will design such a global optimization algorithm to solve PO-PR-SIM in Section VI.

V. Optimal Power Allocation Structure: The Cooperative Relay Case

Comparing CR to PR, the major difference is the presence of the direct link $\mathbf{H}_{sd}$. In Section V-A, we will first study the impact of the direct link and what role it plays in the relationship between PR and CR. Based on this,

we will investigate the structures of optimal values of $\mathbf{A}$ and $\mathbf{Q}$ in Section V-B.

A. The Role of the Direct Link Under Cooperative Relay

Recall that we can write the achievable rate under CR as:

$$I_{CR}(\mathbf{Q},\mathbf{A}) = \frac{1}{2}\log_2\left|\mathbf{I}_{2N_d} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\mathbf{R}^{-1}\right|, \tag{21}$$

where the equivalent end-to-end channel gain matrix $\mathbf{H}$ and noise power $\mathbf{R}$ are defined as

$$\mathbf{H} \triangleq \begin{bmatrix} \bar{\mathbf{H}}_{sd} \\ \mathbf{H}_{sd} \end{bmatrix} \text{ and } \mathbf{R} \triangleq \begin{bmatrix} \bar{\mathbf{R}} & \mathbf{0} \\ \mathbf{0} & \sigma_d^2\mathbf{I}_{N_d} \end{bmatrix}, \tag{22}$$

respectively. We can see from (21) that the value of end-to-end achievable rate under CR is determined by the determinant $\left|\mathbf{I}_{2N_d} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\mathbf{R}^{-1}\right|$. To compute the determinant, we substitute (22) into (21). It then follows that

$$\begin{aligned} &\left|\mathbf{I}_{2N_d} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\mathbf{R}^{-1}\right| \\ &= \left|\mathbf{I}_{2N_d} + \begin{bmatrix} \bar{\mathbf{H}}_{sd} \\ \mathbf{H}_{sd} \end{bmatrix}\mathbf{Q}\begin{bmatrix} \bar{\mathbf{H}}_{sd}^\dagger & \mathbf{H}_{sd}^\dagger \end{bmatrix}\begin{bmatrix} \bar{\mathbf{R}}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_d^2}\mathbf{I}_{N_d} \end{bmatrix}\right| \\ &= \begin{vmatrix} \mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1} & \frac{1}{\sigma_d^2}\bar{\mathbf{H}}_{sd}\mathbf{Q}\mathbf{H}_{sd}^\dagger \\ \mathbf{H}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1} & \mathbf{I}_{N_d} + \frac{1}{\sigma_{sd}^2}\mathbf{H}_{sd}\mathbf{Q}\mathbf{H}_{sd}^\dagger \end{vmatrix}. \end{aligned} \tag{23}$$

Based on the determinant in (23), it can be verified that $I_{CR}(\mathbf{Q},\mathbf{A})$ can be decomposed into two parts as follows:

$$I_{CR}(\mathbf{Q},\mathbf{A}) = \underbrace{\frac{1}{2}\log_2\left|\mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1}\right|}_{\text{Pure AF relay}} + \underbrace{\frac{1}{2}\log_2\left|\mathbf{I}_{N_d} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}(\mathbf{Q}^{-1} + \bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}_{sd})^{-1}\mathbf{H}_{sd}^\dagger\right|}_{\text{Gain from direct link}}. \tag{24}$$

Comparing (24) with (7), we can see that the first term on the right-hand side of (24) is exactly $I_{PR}(\mathbf{Q},\mathbf{A})$. Thus, we can conclude that the second term represents the rate gain due to the cooperation from the direct link. The decomposition in (24) is similar to deriving the mutual information for three-node MIMO AF-relay networks in [12]. We relegate the details of deriving (24) to Appendix C. Expanding $\bar{\mathbf{H}}_{sd}$ and $\bar{\mathbf{R}}$ and noting the block diagonal structure of $\mathbf{A}$, we have the following result.

Proposition 2. *Under CR, the end-to-end achievable data rate gain due to the direct link, denoted by ΔC_{DL}, is given by*

$$\Delta C_{DL} = \frac{1}{2}\log_2\left|\mathbf{I}_{N_d} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}\left(\mathbf{Q}^{-1} + \left(\sum_{i=1}^{M}\mathbf{H}_{rd}^{[i]}\mathbf{A}_i\mathbf{H}_{sr}^{[i]}\right)^\dagger\bar{\mathbf{R}}^{-1}\left(\sum_{i=1}^{M}\mathbf{H}_{rd}^{[i]}\mathbf{A}_i\mathbf{H}_{sr}^{[i]}\right)\right)^{-1}\mathbf{H}_{sd}^\dagger\right|.$$

Remark 4. Proposition 2 can also be understood from another angle. It can be seen that as $\mathbf{H}_{sd} \to \mathbf{0}$ (i.e., the direct link $\mathbf{H}_{sd}$ gets weaker asymptotically), C_{DL} approaches 0. This means that $I_{CR}(\mathbf{Q},\mathbf{A}) \to I_{PR}(\mathbf{Q},\mathbf{A})$ asymptotically, which makes intuitive sense.

We note that, based on (23), $I_{CR}(\mathbf{Q},\mathbf{A})$ can also be decomposed into two parts as follows:

$$I_{CR}(\mathbf{Q},\mathbf{A}) = \underbrace{\frac{1}{2}\log_2\left|\mathbf{I}_{N_d}+\frac{1}{\sigma_d^2}\mathbf{H}_{sd}\mathbf{Q}\mathbf{H}_{sd}^{\dagger}\right|}_{\text{Direct link w/o relays}} + \underbrace{\frac{1}{2}\log_2\left|\mathbf{I}_{N_d}+\bar{\mathbf{H}}_{sd}\left(\mathbf{Q}^{-1}+\frac{1}{\sigma_{sd}^2}\mathbf{H}_{sd}^{\dagger}\mathbf{H}_{sd}\right)^{-1}\bar{\mathbf{H}}_{sd}^{\dagger}\bar{\mathbf{R}}^{-1}\right|}_{\text{Gain from AF relays}}. \quad (25)$$

Upon a closer look at the right-hand side (RHS) of (25), it is readily recognized that the first term is exactly the MIMO capacity expression for the direct link $\mathbf{H}_{sd}$ when the relays are absent. Thus, we can conclude that the second term in the summation represents the rate gain due to the cooperation from the AF relays. The derivation details of the decomposition in (25) is also relegated to Appendix C. After expanding $\bar{\mathbf{H}}_{sd}$ and $\bar{\mathbf{R}}$, we have the following result:

Proposition 3. *Under CR, the end-to-end achievable data rate gain due to the AF relay links, denoted by ΔC_{AF}, is given by*

$$\Delta C_{AF} = \frac{1}{2}\log_2\left|\mathbf{I}_{N_d}+\left(\sum_{i=1}^{M}\mathbf{H}_{rd}^{[i]}\mathbf{A}_i\mathbf{H}_{sr}^{[i]}\right)\left(\mathbf{Q}^{-1}+\frac{1}{\sigma_d^2}\mathbf{H}_{sd}^{\dagger}\mathbf{H}_{sd}\right)^{-1}\left(\sum_{i=1}^{M}\mathbf{H}_{rd}^{[i]}\mathbf{A}_i\mathbf{H}_{sr}^{[i]}\right)^{\dagger}\bar{\mathbf{R}}^{-1}\right|.$$

Remark 5. Similar to Proposition 2, Proposition 3 can also be interpreted in an asymptotic manner. It can be seen that as $\mathbf{H}_{sd}\to\mathbf{0}$, the capacity of the direct link in (25) approaches 0 and $\Delta C_{AF}\to I_{PR}(\mathbf{Q},\mathbf{A})$ asymptotically.

B. The Structure of the Optimal Source and Relay Amplification Matrices

From (25), we can see that once $\mathbf{Q}$ is known, finding an optimal $\mathbf{A}$ is completely determined by maximizing the second term in (25). Also, comparing the second term in (25) to the expression of $I_{PR}(\mathbf{Q},\mathbf{A})$ in (7), we can see that they are of the same structure when $\mathbf{Q}$ is fixed. This means that the optimal structure of $\mathbf{A}$ does not change under CR. We formally state this result in the following theorem.

Theorem 2. *When $\mathbf{Q}$ is fixed, the structural property in (13) continues to hold for $\mathbf{A}$ under CR. However, $\tilde{\mathbf{U}}_{sr}^{[k]}$ in (13) needs to be replaced by the left singular matrix of $\frac{1}{\sigma_r}\mathbf{H}_{sr}(\mathbf{Q}^{-1}+\frac{1}{\sigma_d^2}\mathbf{H}_{sd}^{\dagger}\mathbf{H}_{sd})^{-\frac{1}{2}}$.*

Since $\mathbf{A}$'s optimal structure does not change under CR, the optimal $\mathbf{A}$ can be found by solving an optimization problem similar to PO-PR-SIM. The only difference is that the eigenvalues $\tilde{\lambda}_{sr}^{(i)}$ should be replaced by those of $\frac{1}{\sigma_r}\mathbf{H}_{sr}(\mathbf{Q}^{-1}+\frac{1}{\sigma_d^2}\mathbf{H}_{sd}^{\dagger}\mathbf{H}_{sd})^{-\frac{1}{2}}$. Thus, we omit the formal statement to avoid repetition.

Compared to the PR case, identifying the optimal structure of $\mathbf{Q}$ under CR is slightly more involved. Due to the first term in (25), we need to simultaneously take the direct link $\mathbf{H}_{sd}$ and the equivalent end-to-end relay channel $\bar{\mathbf{H}}_{sd}$ into consideration when finding the optimal power structure of $\mathbf{Q}$. This means that the optimal structure of $\mathbf{Q}$ depends on the value of $\mathbf{A}$. This is different from the PR case, where the optimal structure of $\mathbf{Q}$ is solely dictated by $\mathbf{H}_{sr}$.

To determine the optimal power structure of $\mathbf{Q}$ under CR, we rewrite the achievable rate expression in (8). It can be verified that $I_{CR}(\mathbf{Q},\mathbf{A}) = \frac{1}{2}\log_2\left|\mathbf{I}_{2N_d}+\hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^{\dagger}\right|$, where $\hat{\mathbf{H}}\triangleq\mathbf{R}^{-\frac{1}{2}}\mathbf{H}$. Then, we can see from this expression

that the cooperative relay network is equivalent to a point-to-point MIMO channel with the channel gain matrix being $\hat{\mathbf{H}}$. Thus, from the classical water-filling result [23], we have the following result:

Proposition 4. *The source covariance matrix* $\mathbf{Q}$ *under CR has the following structure at optimality:* $\mathbf{Q} = \hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^{\dagger}$*, where* $\hat{\mathbf{V}}$ *represents the right singular matrix of* $\hat{\mathbf{H}}$*, i.e.,* $\hat{\mathbf{H}} = \hat{\mathbf{U}}\mathbf{\Sigma}_{sr}\hat{\mathbf{V}}^{\dagger}$*, where* $\mathbf{\Sigma}_{sr}$ *is a diagonal matrix, and* $\hat{\mathbf{U}}$ *and* $\hat{\mathbf{V}}$ *are the associated left and right singular matrices, respectively.*

Remark 6. Theorem 2 and Proposition 4 indicate that, even though the channels in the CR case become more complex, the same matching structure of an optimal $\mathbf{A}$ matrix continues to hold. The matrix $\mathbf{Q}$ also exhibits a matching structure at optimality, although the matching has to jointly consider $\mathbf{H}_{sr}$, $\mathbf{H}_{rd}$, and $\mathbf{H}_{sd}$.

VI. Optimization Algorithm

Based on the structural results in the previous two sections, we now study how to determine an optimal $\mathbf{Q}$ and $\mathbf{A}$. Since Theorem 2 and Proposition 4 say that the optimal structures of $\mathbf{A}$ and $\mathbf{Q}$ are the same under PR and CR, it suffices to only consider the design of an optimization algorithm for the PR case.

Recall that PO-PR can be decomposed into an inner subproblem with respect to $\mathbf{A}$ where $\mathbf{Q}$ is held fixed, and an outer main program with respect to $\mathbf{Q}$. We first examine the convexity property of the outer subproblem. It can be seen from PO-PR-SIM that the objective function is concave with respect to the $\tilde{\lambda}_j^{(i)}$-variables. Thus, if we know how to optimize $\mathbf{A}$ for each given $\mathbf{Q}$, then an optimal $\mathbf{Q}$ can be solved by any standard convex optimization tools. Therefore, we would only focus on the inner subproblem in the remainder of this section.

A. Convexity Property of PO-PR-SIM

For convenience, we repeat the inner subproblem of PO-PR-SIM below.

PO-PR-SIM:

$$\begin{aligned}
\text{Maximize } & \frac{1}{2}\sum_{j=1}^{D}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}\delta_j^2}{\lambda_{rd}^{(j)}\delta_j^2+\frac{\sigma_d^2}{\sigma_r^2}}\right] \\
\text{subject to } & \sum_{j=1}^{MN_r}(\mathbf{S}_i)_{jj}(d_j^{(i)})^2+\sum_{j=1}^{MN_r}\sum_{k=1,k\neq j}^{MN_r}(\mathbf{S}_i)_{jk}d_j^{(i)}d_k^{(i)}\leq P_R, \qquad i=1,2,\ldots,M \\
& \delta_j-\sum_{i=1}^{M}d_j^{(i)}=0, \qquad j=1,2,\ldots,N_r.
\end{aligned}$$

To solve the inner subproblem (i.e., with $\tilde{\lambda}_{sr}^{(j)}$ being fixed), we again examine its convexity property first. We note that in PO-PR-SIM, the second constraint is linear. The first constraint is in a Schur product form involving a principal of a positive semidefinite matrix, which means it's convex as well. Thus, the problem is a convex program if its objective is concave. However, the objective function of PO-PR-SIM can be written as a difference of two concave functions:

$$\frac{1}{2}\sum_{j=1}^{D}\log_2\left((1+\tilde{\lambda}_{sr}^{(j)})\lambda_{rd}^{(j)}\delta_j^2+\frac{\sigma_d^2}{\sigma_r^2}\right)-\frac{1}{2}\sum_{j=1}^{D}\log_2\left(\lambda_{rd}^{(j)}\delta_j^2+\frac{\sigma_d^2}{\sigma_r^2}\right),$$

which implies that it is not concave.

In fact, by checking the second order condition (i.e., the positive definiteness of the Hessian matrix) of the objective function of PO-PR-SIM, we can show that PO-PR-SIM is convex when the diagonal elements $d_j^{(i)}$ (constrained by relay power P_R) and channels $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ satisfy the following conditions:

$$\left(\sum_{i=1}^{M} d_j^{(i)}\right)^2 \geq \frac{\sigma_d^2}{6\lambda_{rd}^{(j)}\sigma_c^2}\left[\sqrt{\left(\frac{\tilde{\lambda}_{sr}^{(j)}+2}{\tilde{\lambda}_{sr}^{(j)}+1}\right)^2+\left(\frac{12}{\tilde{\lambda}_{sr}^{(j)}+1}\right)}-\left(\frac{\tilde{\lambda}_{sr}^{(j)}+2}{\tilde{\lambda}_{sr}^{(j)}+1}\right)\right],\ j=1,\ldots,D. \tag{26}$$

The derivation of (26) is similar to that in [18] and is thus omitted for brevity. By observing (26), we can see that as $\tilde{\lambda}_{sr}^{(j)} \to \infty$ or $\lambda_{rd}^{(j)} \to \infty$, the RHS approaches 0. Therefore, we have the following result.

Proposition 5. *The threshold value on* $\left(\sum_{i=1}^{M} d_j^{(i)}\right)^2$ *approaches to zero as* $\tilde{\lambda}_{sr}^{(j)} \to \infty$ *or* $\lambda_{rd}^{(j)} \to \infty$.

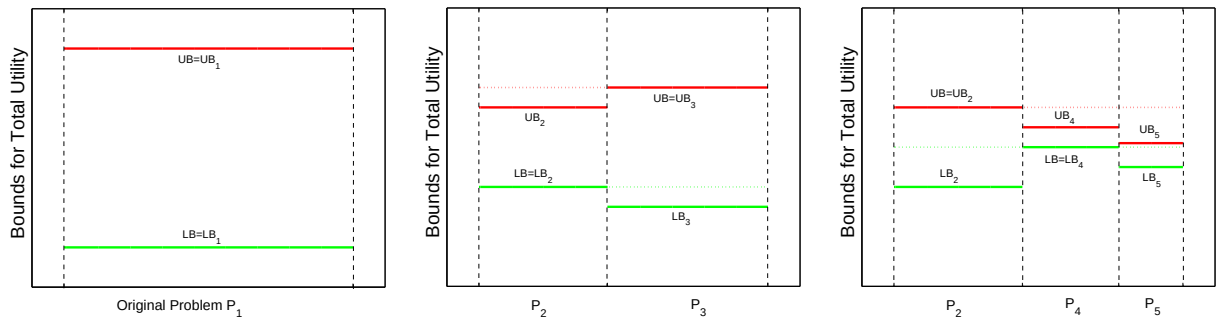
Remark 7. The above proposition implies that it is easier for PO-PR-SIM to be convex when $\tilde{\lambda}_{sr}^{(j)}$ or $\lambda_{rd}^{(j)}$ is relatively large. Physically, $\tilde{\lambda}_{sr}^{(j)}$ and $\lambda_{rd}^{(j)}$ represent the quality of the source-relay and relay-destination channels. Proposition 5 says that PO-PR-SIM is more likely to be convex if the channels $\mathbf{H}_{sr}$ and $\mathbf{H}_{rd}$ are strong.

Due to the nonconvexity of PO-PR-SIM, designing algorithms to determine its global optimal solution remains challenging even though PO-PR-SIM is in scalar form. Rather than settling with heuristic or local optimization algorithms, we propose a global optimization approach based on the branch-and-bound global optimization method coupled with a convex hull relaxation (BB/CHR) specifically tailored to PO-PR-SIM, which exploits the special structure in PO-PR-SIM. To the best of our knowledge, our work is the first that considers solving nonconvex MIMO-CN problems using a global optimization approach.

B. A Global Optimization Approach

Our proposed BB/CHR approach starts by further reformulating the inner PO-PR-SIM problem. Before doing so, we briefly describe the basic idea of our general BB/CHR framework. Then, we will show how to identify the special structure in PO-PR-SIM to facilitate the derivation of an appropriate CHR.

BB/CHR: Basic Idea. For a general nonconvex optimization problem, conventional convex programming methods (e.g., interior methods [24]) can at best yield local optimal solutions. In contrast, our proposed BB/CHR procedure in this paper is a powerful global optimization approach that guarantees finding a global optimal solution [25]–[27]. The basic idea of BB/CHR is that we can construct a higher dimensional convex-hull relaxation (CHR) for the original nonconvex problem, which can be used to efficiently compute a global upper bound, UB, for the original nonconvex problem. The solution to the CHR is either a feasible solution to the original nonconvex problem or, if not feasible, can be used as a starting point for a local search algorithm to find a feasible solution to the original nonconvex problem. This feasible solution will then serve to provide a global lower bound, LB, and an incumbent solution to the original nonconvex problem. However, it is worth pointing out that local search is not necessary in this paper since the CHR for our problem is still subject to the same power constraint of the original problem. As a result, solving the CP relaxation of our problem still gives us a feasible $d_j^{(i)}$-solution to the original



(a) Iteration 1 (b) Iteration 2 (c) Iteration 3

Fig. 5. Illustration of branch-and-bound solution procedure.

problem. This will become clearer later after we introduce the CHR. The proposed branch-and-bound process then proceeds by tightening UB and LB, and terminates when $LB \geq (1-\epsilon)UB$ is satisfied, where ϵ is some desired termination criterion. There is a formal proof that the general BB process converges to a global optimal solution as long as the partitioning intervals are compact (see, e.g., [25]–[27]), which obviously holds in our PO-PR-SIM.

The detailed branch-and-bound process works as follows. First, we substitute single new variables to represent terms causing nonconvexity in the problem. Define the relaxation error for a nonconvex term in the problem as the difference between the value of this term and the value of its corresponding new variable in the relaxation solution. If such relaxation errors are significant, then the lower bound LB can be far away from the upper bound UB. To close this gap, we must construct a tight CHR, i.e., having smaller relaxation errors. This can be achieved, as described below, by further narrowing the bounding interval length of the partitioning variables (more on this later). Specifically, the branch-and-bound algorithm selects a partitioning variable having the maximum relaxation error and divides its bounding interval into two intervals by cutting it at the relaxation solution. Then the original problem having the greatest upper bound $UB_1 = UB$ (denoted as Problem P_1 as shown in Fig. 5(a)) is divided into two new subproblems (denoted as Problems P_2 and P_3 as shown in Fig. 5(b)). Again, we solve the CHR for these two new problems, and thereby obtain LB_2 and UB_2 for P_2, and LB_3 and UB_3 for P_3. Since the relaxations for P_2 and P_3 are both tighter than that in Problem P_1, we must have $\max\{UB_2, UB_3\} \leq UB_1$. The upper bound of the original problem is updated from $UB = UB_1$ to $UB = \max\{UB_2, UB_3\}$ and the lower bound of the original problem is updated from LB to $LB = \max\{LB, LB_2, LB_3\}$, with the incumbent solution being updated in case an improvement results. As a result, we now have a smaller gap between UB and LB. If $LB \geq (1-\epsilon)UB$, the branch-and-bound process is terminated. Otherwise, we choose a subproblem that has the maximum upper bound (Problem P_3 in Fig. 5(b)) and further perform a partitioning for this problem as above. In the next iteration, as shown in Fig. 5(c), P_3 is partitioned into P_4 and P_5. Upon solving the CHR for P_4 and P_5, respectively, we have two sets of upper and lower bounds (LB_4, UB_4) and (LB_5, UB_5). Notice that in Problem P_5 in the illustration in Fig. 5(c), $(1-\epsilon)UB_5 < LB$. This means that further partition on P_5 is unnecessary. As such, we can discard P_5 from the problem list. After we finish partitioning P_3, the global upper bound would be changed

Algorithm 1 BB/CHR Solution Procedure

Initialization:

1. Let optimal solution $\psi^* = \emptyset$. The initial lower bound $LB = -\infty$.
2. Determine partitioning variables (variables associated with CHR) and derive their initial bounding intervals.
3. Let the initial problem list contain only the original problem, denoted by P_1.
4. Modify constraints for partitioning variables to build a CHR. Denote the solution to CHR as $\hat{\psi}_1$ and its objective value as the upper bound UB_1.

Main Loop:

5. Select problem P_z that has the largest upper bound among all problems in the problem list.
6. Find, if necessary, a feasible solution ψ_z via a local search algorithm for Problem P_z. Denote the objective value of ψ_z by LB_z.
7. If $LB_z > LB$ then let $\psi^* = \psi_z$ and $LB = LB_z$. If $LB \geq (1-\epsilon)UB$ then stop with the $(1-\epsilon)$-optimal solution ψ^*; else, remove all problems $P_{z'}$ having $(1-\epsilon)UB_{z'} \leq LB$ from the problem list.
8. Compute relaxation error for each partitioning variable.
9. Select a partitioning variable having the maximum relaxation error and divide its bounding interval into two new intervals by partitioning at its value in $\hat{\psi}_z$.
10. Remove the selected problem P_z from the problem list, construct two new problems P_{z1} and P_{z2} based on the two partitioned intervals.
11. Compute two new upper bounds UB_{z1} and UB_{z2} by solving the CHR of P_{z1} and P_{z2}, respectively.
12. If $LB < (1-\epsilon)UB_{z1}$ then add problem P_{z1} to the problem list. If $LB < (1-\epsilon)UB_{z2}$ then add problem P_{z2} to the problem list.
13. If the problem list is empty, stop with the $(1-\epsilon)$-optimal solution ψ^*. Otherwise, go to Step 5 again.

to $UB = \max\{UB_2, UB_4, UB_5\}$, which is UB_2 in this case. The general framework of BB/CHR is summarized in Algorithm 1.

In the remainder of this section, we will develop the key components in the BB/CHR framework, which are problem-specific.

Objective function reformulation. Observe that in PO-PR-SIM, nonconvexity only appears in the objective function. Therefore, we start by rewriting the objective function of the inner problem as

$$\frac{1}{2}\sum_{j=1}^{D}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}g_j}{\lambda_{rd}^{(j)}g_j+\sigma^2}\right], \tag{27}$$

where $\sigma^2 \triangleq \frac{\sigma_d^2}{\sigma_r^2}$ and $g_j \triangleq \delta_j^2 \geq 0$. Note that, although not obvious, (27) is convex with respect to $g_j \geq 0$, $\forall j = 1, \ldots, N$. This can be readily verified by checking the second order condition of (27). With this change of variables, the inner problem can be written as follows:

PO-PR-SIM:

$$\begin{aligned}
&\text{Maximize } \frac{1}{2}\sum_{j=1}^{D}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}g_j}{\lambda_{rd}^{(j)}g_j+\sigma^2}\right] \\
&\text{subject to } g_j = \delta_j^2 && \forall j = 1, \ldots, D \\
&\quad \sum_{j=1}^{MN_r}(\mathbf{S}_i)_{jj}(d_j^{(i)})^2 + \sum_{j=1}^{MN_r}\sum_{k=1,k\neq j}^{MN_r}(\mathbf{S}_i)_{jk}d_j^{(i)}d_k^{(i)} \leq P_R, && \forall i = 1, \ldots, M \\
&\quad \delta_j - \sum_{i=1}^{M} d_j^{(i)} = 0, && \forall j = 1, \ldots, D.
\end{aligned}$$

Although this reformulated problem remains nonconvex (since the new constraints $g_j = \delta_j^2$, $\forall j = 1, \ldots, N_r$ are nonconvex), we have shifted the nonconvex part from the objective function to the constraints. As we shall see immediately, with this reformulation, we can apply a powerful technique called the *Reformulation-Linearization Technique* (RLT) [25], [26].

Convexifying the quadratic equality constraints using RLT. The basic idea in our reformulation is to approximate the quadratic equality constraint $g_j = \delta_j^2$ using its convex hull relaxation, which by definition is the tightest possible convex relaxation. We first observe that if we replace $g_j = \delta_j^2$ by $g_j \geq \delta_j^2$, then the latter is convex. However, since PO-PR-SIM is a maximization problem in g_j, we need an additional constraint to bound g_j from above because otherwise, PO-PR-SIM will be unbounded. To this end, we can apply RLT as follows. First, we note that

$$\delta_j^L \leq \delta_j \leq \delta_j^U, \tag{28}$$

where δ_j^L and δ_j^U are some appropriate lower and upper bounds for δ_j, respectively (more on the appropriate values of δ_j^L and δ_j^U later). From (28), we derive the following so-called bounding-factor constraint:

$$(\delta_j - \delta_j^L)(\delta_j - \delta_j^U) \leq 0, \tag{29}$$

which, upon using the substitution $g_j = \delta_j^2$, yields:

$$g_j - (\delta_j^L + \delta_j^U)\delta_j + \delta_j^L \delta_j^U \leq 0. \tag{30}$$

Observe now that the original non-convex constraint $g_j = \delta_j^2$ has been relaxed into two constraints in g_j and δ_j:

$$\delta_j^2 - g_j \leq 0 \qquad \forall j = 1, \ldots, D, \tag{31}$$

$$g_j - (\delta_j^L + \delta_j^U)\delta_j + \delta_j^L \delta_j^U \leq 0 \qquad \forall j = 1, \ldots, D. \tag{32}$$

Geometrically, the reformulated constraint in (31) and the RLT constraint in (32) represent the convex hull of the set defined by $g_j = \delta_j^2$, $\delta_j \in [\delta_j^L,\ \delta_j^U]$, as shown in Fig. 6. When the bounds δ_j^L and δ_j^U are far from each other (i.e., the interval $[\delta_j^L, \delta_j^U]$ is large), the convex hull relaxation may not be a tight approximation of the quadratic curve. However, as δ_j^L and δ_j^U get tighter, as shown in Fig. 7, we can see that the convex hull becomes an excellent approximation of the quadratic curve.

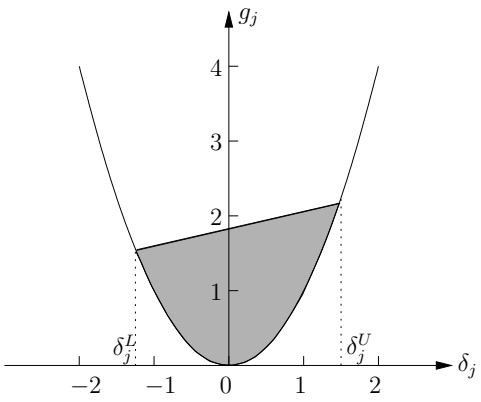


Fig. 6. The quadratic curve $g_j = \delta_j^2$ is approximated by a polyhedron formed by the linear RLT constraints.

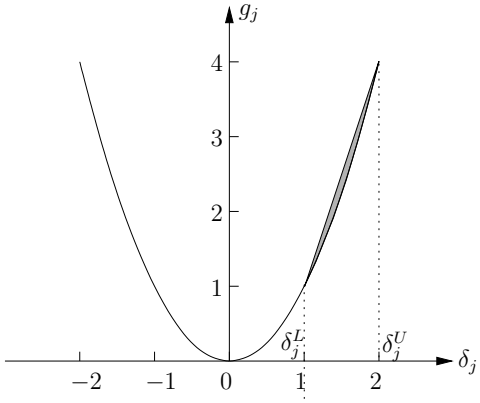


Fig. 7. The quadratic curve $g_j = \delta_j^2$ can be well approximated by the linear RLT polyhedron when the interval $[\delta_j^L, \delta_j^U]$ is small.

After the above convex relaxation, we have the following CHR of inner problem:

R-PO-PR-SIM:

$$\begin{aligned}
\text{Maximize } & \frac{1}{2}\sum_{j=1}^{D}\log_2\left[1+\frac{\tilde{\lambda}_{sr}^{(j)}\lambda_{rd}^{(j)}g_j}{\lambda_{rd}^{(j)}g_j+\sigma^2}\right] \\
\text{subject to } & \delta_j^2 - g_j \le 0 && \forall j=1,\ldots,D \\
& g_j - (\delta_j^L+\delta_j^U)\delta_j + \delta_j^L\delta_j^U \le 0 && \forall j=1,\ldots,D \\
& \sum_{j=1}^{MN_r}(\mathbf{S}_i)_{jj}(d_j^{(i)})^2 + \sum_{j=1}^{MN_r}\sum_{k=1,k\neq j}^{MN_r}(\mathbf{S}_i)_{jk}d_j^{(i)}d_k^{(i)} \le P_R, && \forall i=1,\ldots,M \\
& \delta_j - \sum_{i=1}^{M} d_j^{(i)} = 0, && \forall j=1,\ldots,D.
\end{aligned}$$

One important remark on R-PO-PR-SIM is in order: if PO-PR-SIM happens to be a convex problem (see Proposition 5), the branch-and-bound process in our proposed BB/CHR will also converge in *one* step. This is due to the use of convex hull relaxation, which *automatically* yields the optimal solution to the original problem. This means that our BB/CHR method does *not* incur any extra cost if R-PO-PR-SIM is well-conditioned. This interesting fact will also be seen in numerical results presented later.

Partitioning variables and their upper and lower bounds The partitioning variables in the branch-and-bound (BB) process are those that are involved in nonconvex terms, for which we have therefore defined new variables, and whose bounding intervals will need to be partitioned during the branch-and-bound algorithm [25]–[27]. In R-PO-PR-SIM, these partitioning variables are δ_j, $j=1,\ldots,D$. For these partitioning variables, we need to derive tight upper and lower bounds for δ_j^U and δ_j^L, respectively.

To this end, we let $\bar{\mathbf{S}}_i^{(j)} \in \mathbb{R}^{(MN_r-1)\times MN_r}$ be the matrix obtained by taking the real parts of $\mathbf{S}_i$ and then deleting the j-th row. Then, let $\bar{\mathbf{v}}_i^{(j)} = [\bar{v}_{i,1}^{(j)} \ \cdots \ \bar{v}_{i,MN_r}^{(j)}]^T$ be the right singular vector corresponding to the zero singular value of $\bar{\mathbf{S}}_i^{(j)}$. Further, define $\hat{\mathbf{v}}_i^{(j)}$ as the scaled version of a right singular vector of $\bar{\mathbf{v}}_i^{(j)}$ obtained by

$\hat{\mathbf{v}}_i^{(j)} = \frac{1}{\bar{v}_{i,1}^{(j)}} \bar{\mathbf{v}}_i^{(j)} = \left[1 \ \frac{\bar{v}_{i,2}^{(j)}}{\bar{v}_{i,1}^{(j)}} \cdots \frac{\bar{v}_{i,MN_T}^{(j)}}{\bar{v}_{i,1}^{(j)}} \right]^T$. Then, we have the following result:

Lemma 4. *The constraint* $\sum_{j=1}^{MN_r} (\mathbf{S}_i)_{jj} (d_j^{(i)})^2 + \sum_{j=1}^{MN_r} \sum_{k=1,k\neq j}^{MN_r} (\mathbf{S}_i)_{jk} d_j^{(i)} d_k^{(i)} \leq P_R$ *implies the following inequality for each* $d_j^{(i)}$*:*

$$|d_j^{(i)}| \leq \sqrt{\frac{P_R}{(\hat{\mathbf{v}}_i^{(j)})^\dagger \mathbf{S}_i \hat{\mathbf{v}}_i^{(j)}}}, \quad \forall i = 1, \ldots, M, j = 1, \ldots, MN_r. \tag{33}$$

Lemma 4 can be proved by considering the following two optimization problems:

$$\begin{array}{ll} \text{Maximize} & d_j^{(i)} \\ \text{subject to} & \sum_{j=1}^{MN_r} (\mathbf{S}_i)_{jj} (d_j^{(i)})^2 + \sum_{j=1}^{MN_r} \sum_{k=1,k\neq j}^{MN_r} (\mathbf{S}_i)_{jk} d_j^{(i)} d_k^{(i)} \leq P_R. \end{array} \tag{34}$$

and

$$\begin{array}{ll} \text{Minimize} & d_j^{(i)} \\ \text{subject to} & \sum_{j=1}^{MN_r} (\mathbf{S}_i)_{jj} (d_j^{(i)})^2 + \sum_{j=1}^{MN_r} \sum_{k=1,k\neq j}^{MN_r} (\mathbf{S}_i)_{jk} d_j^{(i)} d_k^{(i)} \leq P_R. \end{array} \tag{35}$$

Since these two optimization problems are convex and the Slater condition obviously holds, the KKT optimality conditions are both necessary and sufficient for Problems (34) and (35) [24]. Then, the result in (33) follows from analyzing the KKT system of (34) and (35). We relegate the proof details of Lemma 4 to Appendix D.

Next, observe that $\delta_j^U = \max \delta_j = \max \sum_{i=1}^M d_j^{(i)} = \sum_{i=1}^M \max d_j^{(i)}$ and $\delta_j^L = \min \delta_j = \min \sum_{i=1}^M d_j^{(i)} = \sum_{i=1}^M \min d_j^{(i)}$. Then, from Lemma 4, the following result is straightforward:

Lemma 5. *The upper and lower bounds of* δ_j *can be respectively computed as*

$$\delta_j^L = -\sum_{i=1}^M \sqrt{\frac{P_R}{(\hat{\mathbf{v}}_i^{(j)})^\dagger \mathbf{S}_i \hat{\mathbf{v}}_i^{(j)}}}, \quad \textit{and} \quad \delta_j^U = \sum_{i=1}^M \sqrt{\frac{P_R}{(\hat{\mathbf{v}}_i^{(j)})^\dagger \mathbf{S}_i \hat{\mathbf{v}}_i^{(j)}}}. \tag{36}$$

C. Numerical Results

We present some numerical results to show the efficacy of our proposed BB/CHR method. First, we consider a network shown in Fig. 8, which satisfies the convexity condition in (26). In this example, the source node is N3 and the destination node is N4. Each node in the network is equipped with four antennas. The transmit power levels are chosen as $P_T = P_R = 1$W, which are large enough so that the convexity condition (26) is satisfied. The path-loss index is chosen to be 3. The BB termination criterion ϵ is 10^{-6}. The BB convergence process of the global upper bound and the global lower bound (the incumbent optimal and feasible solution) is shown in Table I. It can be seen that the gap between the global upper and lower bounds is already negligible after solving the convex relaxation R-PO-PR-SIM for the original problem, meaning that the branch-and-bound process converges in one step for this case. This confirms our expectation that BB/CHR does not incur any extra cost when the original problem happens to be convex.

On the other hand, consider another network example as shown in Fig. 9. In this example, the source node is N1 and the destination node is N4. Again, each node in the network is equipped with four antennas. However, the transmit power levels are chosen as $P_T = P_R = 100$ mW, which are weak enough so that the convexity condition in (26) is violated. The path-loss index is again chosen to be 3 and the BB termination criterion ϵ is set to 10^{-6} as

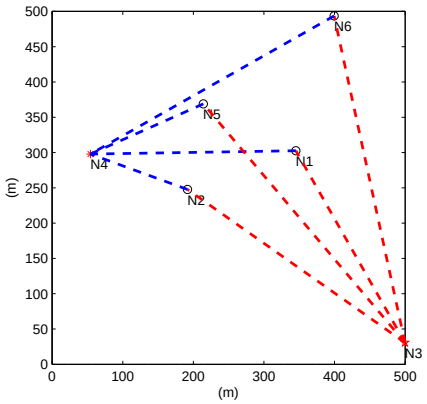


Fig. 8. A four-relay network example satisfying the convexity condition in (26).

TABLE I

CONVERGENCE PROCESS OF THE GLOBAL UPPER AND LOWER BOUNDS OF THE OBJECTIVE VALUE (IN BITS/S/HZ) WHEN PO-PR-SIM IS CONVEX.

No. Iterations	Global upper bound	Global lower bound
1	16.9597	16.9595

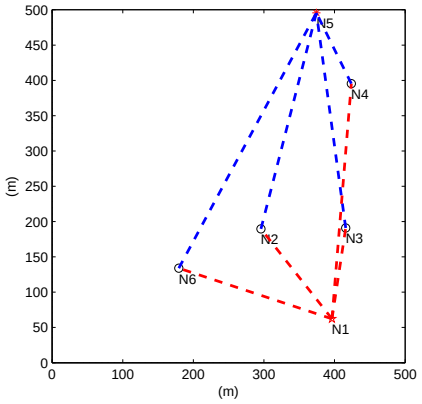


Fig. 9. A four-relay network example violating the convexity condition in (26).

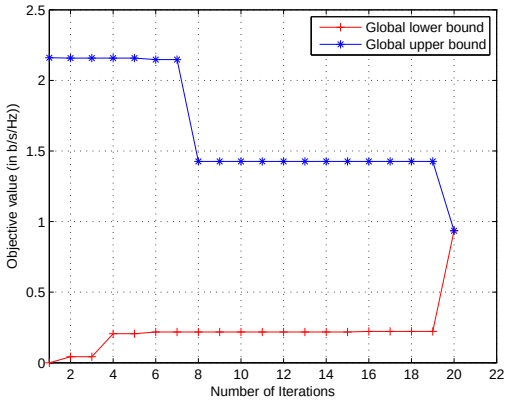


Fig. 10. The convergence process of BB/CHR when PO-PR-SIM is nonconvex.

before. The BB convergence process of the global upper bound and the global lower bound (the incumbent optimal and feasible solution) is shown in Fig. 10. It can be seen that it only takes 20 iterations of branch-and-bound to find a global optimal solution. This shows the efficacy and efficiency of our proposed BB/CHR method for nonconvex cases of PO-PR-SIM.

Next, we study the scaling of iterations as the number of relay nodes and the number of antennas per node grow. In general, the number of branchings in the branch-and-bound framework depends on the number of partitioning variables, which is the number of δ-variables in this paper. From previous discussions on BB/CHR, we can see that the number of δ-variables is determined by the effective degree of freedom D, which is in turn determined by the number of antennas per node. Thus, we expect that the number of antennas should have a more direct impact on the number of branching iterations than the number of relay nodes does. We first fix the number of relay nodes to 4 and vary the number of antennas per node from 2 to 6, which is the practical range of the number of antennas in real systems. The transmit power of each node is set to 100 mW and all other simulation settings

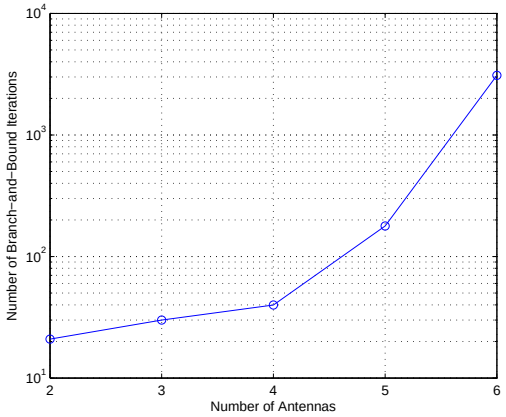


Fig. 11. The scaling of the number of branching iterations with respect to the number of antennas per node.

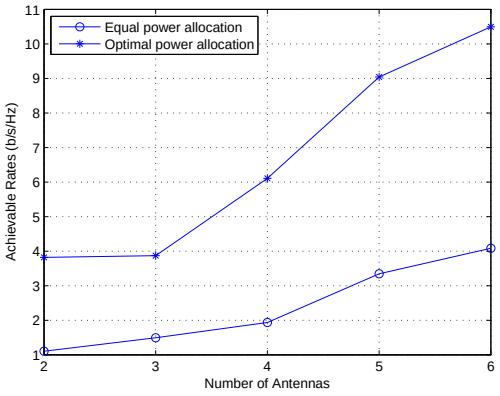


Fig. 12. The achievable rate comparison between equal power allocation and optimal power allocation as the number of antennas per node changes from 2 to 6.

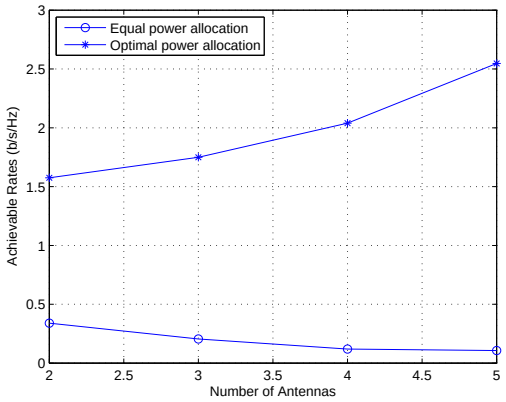


Fig. 13. The achievable rate comparison between equal power allocation and optimal power allocation as the number of relay nodes changes from 2 to 5.

remain the same as in the previous two examples. In this case, the scaling of the number of branch-and-bound iterations is shown in Fig. 11 (in log-scale). Each data point in Fig. 11 is averaged over 50 randomly generated network examples. It can seen that the averaged number of iterations changes from 21 to 3090, which increases roughly exponentially. This confirms that the number of branching iterations depends on the number of antennas per node. The exponential increase of iterations is an expected phenomenon when searching for global optimal solutions for NP-hard nonconvex optimization problems. However, we note that for the practical numbers of antennas, the branch-and-bound process converges reasonably fast. On the other hand, we also evaluate the scaling of the number of branching iterations with respect to the number of relay nodes. Again, for each setting, the result is averaged over 50 randomly generated network examples. We fix the antenna per node to be 2 and vary the number of relays from 2 to 6. The results obtained are 16, 25, 20, 18, and 19, respectively, which stay roughly in the same range and are not sensitive to the number of relays.

It is also interesting to study how much rate gain can be obtained through an optimal power allocation. In

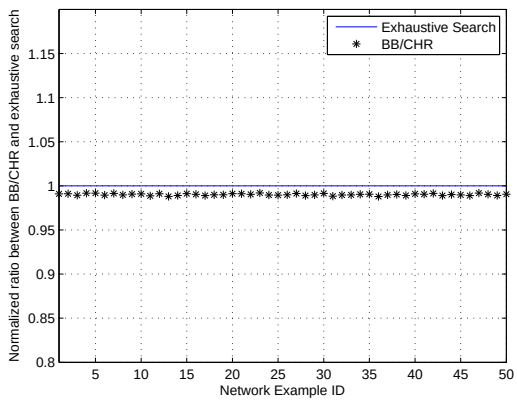


Fig. 14. The comparison between BB/CHR algorithm and the exhaustive search method.

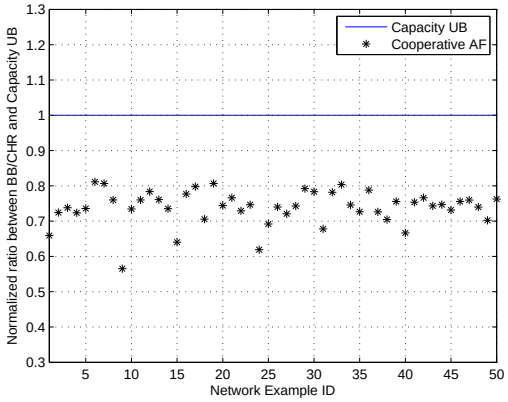


Fig. 15. The achievable rate comparison between equal power allocation and optimal power allocation.

Fig. 12, we compare the achievable rates under optimal power allocation and equal power allocation as the number of antennas per node increases. It can be seen that the achievable rate gain is significant after an optimal power allocation. For these five settings, the average ratios between an equal power allocation and an optimal power allocation are 0.2849, 0.3914, 0.3171, 0.3699, and 0.3891, respectively. In other words, the achievable rates under equal power allocation are no more than 40% of that of the optimal power allocation. In Fig. 13, we compare the achievable rates under optimal power allocation and equal power allocation as the number of relay nodes increases. Again, we can see that the achievable rate gain is significant after optimal power allocation. For these four settings, the average ratios between equal power allocation and optimal power allocations are 0.2154, 0.1173, 0.0586, and 0.0419. respectively. We note that, after optimal power allocation, the achievable rates increases as the number of relay nodes grows. However, under a simple equal power allocation scheme, the achievable rate performance actually decreases. This highlights the importance of power allocation optimization in multi-relay MIMO-CN.

Next, we compare the achievable rates between branch-and-bound and exhaustive search to verify that the branch-and-bound indeed achieves a global optimal solution. Since exhaustive search quickly becomes intractable as the problem sizes gets larger, we choose a two-relay network with two antennas per node as our base setting for exhaustive search. The exhaustive search is conducted as follows. We then quantize each interval and test all possible combinations of quantized values of δ_j. The combination with the largest objective value is recorded. We tested 50 randomly generated network examples. The results are shown in Fig. 14, where we normalize the solutions of the BB/CHR with respect to those of the exhaustive search. It can be seen that the normalized ratios are close to one in all cases. Thus, we can conclude that, within the accuracy of numerical computation, the BB/CHR algorithm achieves a global optimal solution.

Finally, since the capacity of multi-relay cooperative channel remains an open problem, it is instructive to study the maximum achievable rate of AF-based cooperative schemes compared with a capacity upper bound of the cooperative relay network. Here, we use the minimum cut-set bound as a capacity upper bound (i.e., the minimum of the MIMO broadcast channel formed by the source-relay links and the MIMO multiple access channel formed by

the relay-destination links). We tested 50 randomly generated network examples with four relays and four antennas. The results are shown in Fig. 15, where we normalize the results of the AF-based cooperative scheme with respect to the upper bound. We can see from Fig. 15 that most of the normalized results fall in the range between 60% and 80%. The mean of these 50 examples are 73.9%, implying that, on average, the AF-based cooperative scheme achieves at least 73.9% of the true capacity in these 50 examples.

VII. Conclusion

In this paper, we have investigated the structural properties of optimal power allocation in MIMO-CN and constructed a tractable optimization framework that is amenable for networking analysis. Our contributions include generalizing the matching result under the basic three-node setting to the multi-relay setting, for which the relay power allocation has been heretofore unknown. We also quantified the performance gain due to cooperative relay and established the connection between cooperative relay and pure relay. Based on the derived structural insights, we reduced the MIMO-CN rate maximization problem to an equivalent scalar form, which allowed us to develop an efficient global optimization algorithm by using branch-and-bound coupled with a custom designed convex hull relaxation for the simplified problem. Our results in this paper offer important analytical tools and insights to fully exploit the potential of AF-based MIMO-CN with multiple relay nodes. More importantly, our results enable future large-scale cooperative network research to go beyond the basic three-node relay setting.

Appendix A
Comparison between AF and DF under the Multi-Relay Setting

In this appendix, we compare the end-to-end achievable rate performance between AF and DF under the multi-relay setting as shown in Fig. 2. For simplicity, we do not consider fading in this example. As shown in Fig. 2, the source node S wants to communicate with the destination node D, possibly via the help of M *parallel* relay nodes $R_1, R_2, ..., R_M$. All communications share the same frequency band. Due to self-interference, the relay nodes cannot send and receive at the same time. Thus, a unit time-frame is divided into two time slots and the relay nodes operate in half-duplex mode.

In both AF and DF, we consider pure relay (PR) and cooperative relay (CR) paradigms. Note that in Fig. 2, since there are M copies of the same message available from the relays, the destination can combine them to increase the received SNR. Here, we assume that the destination employs maximum-ratio-combining (MRC) to maximize the received SNR.

A. Pure Relay (PR)

In PR with MRC, the end-to-end mutual information under AF can be computed as

$$I_{AF}^{PR} = \frac{1}{2}\log_2\left(1+\sum_{i=1}^{M}\frac{\mathsf{SNR}_{sr_i}\mathsf{SNR}_{r_id}}{1+\mathsf{SNR}_{sr_i}+\mathsf{SNR}_{r_id}}\right), \tag{37}$$

where SNR_{sr_i} and SNR_{r_id} represent the signal-to-noise ratios for the source-ith relay and ith relay-destination links, respectively.

On the other hand, the end-to-end mutual information of DF under PR can be written as

$$I_{DF}^{PR} = \min\left\{\min_{i=1,\ldots,M}\left\{\frac{1}{2}\log_2\left(1+\mathsf{SNR}_{sr_i}\right)\right\}, \frac{1}{2}\log_2\left(1+\sum_{i=1}^{M}\mathsf{SNR}_{r_id}\right)\right\}. \tag{38}$$

From (37) and (38), it is easy to see that the comparison between I_{AF}^{PR} and I_{AF}^{CC} depends on the relationship between the effective SNRs $\sum_{i=1}^{M}\frac{\mathsf{SNR}_{sr_i}\mathsf{SNR}_{r_id}}{1+\mathsf{SNR}_{sr_i}+\mathsf{SNR}_{r_id}}$ and $\min\{\mathsf{SNR}_{sr_1},\ldots,\mathsf{SNR}_{sr_M},\sum_{i=1}^{M}\mathsf{SNR}_{r_id}\}$. However, the relationship between $\sum_{i=1}^{M}\frac{\mathsf{SNR}_{sr_i}\mathsf{SNR}_{r_id}}{1+\mathsf{SNR}_{sr_i}+\mathsf{SNR}_{r_id}}$ and $\min\{\mathsf{SNR}_{sr_1},\ldots,\mathsf{SNR}_{sr_M},\sum_{i=1}^{M}\mathsf{SNR}_{r_id}\}$ is *indefinite*. This can be shown by the following simple numerical example:

Example A.1. Consider a homogeneous network where $\mathsf{SNR}_{sr_i} = \mathsf{SNR}_{r_id} = 1$ for all $i = 1,\ldots,M$. It follows from (37) and (38) that the effective SNR for AF is $\frac{M}{3}$ and the effective SNR for DF remains 1. That is, as the number of relays increases, the effective SNR under AF scales linearly, while the effective SNR under DF is held fixed. In this case, it is clear that if $M > 3$, AF outperforms DF. Otherwise, if $M \leq 3$, AF has no advantage over DF.

B. Cooperative Relay (CR)

Under CR with MRC, the destination node can further coherently combine the signal from the direct link with the signals from the relays. Thus, we have the end-to-end mutual information under AF as

$$I_{AF}^{CR} = \frac{1}{2}\log_2\left(1+\mathsf{SNR}_{sd}+\sum_{i=1}^{M}\frac{\mathsf{SNR}_{sr_i}\times\mathsf{SNR}_{r_id}}{\mathsf{SNR}_{sr_i}+\mathsf{SNR}_{r_id}+1}\right). \tag{39}$$

On the other hand, the end-to-end mutual information of DF under CR can be written as

$$I_{DF}^{CR} = \min\left\{\min_{i=1,\ldots,M}\left\{\frac{1}{2}\log_2\left(1+\mathsf{SNR}_{sr_i}\right)\right\}, \frac{1}{2}\log_2\left(1+\mathsf{SNR}_{sd}+\sum_{i=1}^{M}\mathsf{SNR}_{r_id}\right)\right\}. \tag{40}$$

Again, we can see that the comparison between I_{AF}^{CR} and I_{DF}^{CR} depends on the relationship between the effective SNRs $\mathsf{SNR}_{sd}+\sum_{i=1}^{M}\frac{\mathsf{SNR}_{sr_i}\times\mathsf{SNR}_{r_id}}{\mathsf{SNR}_{sr_i}+\mathsf{SNR}_{r_id}+1}$ and $\min\{\mathsf{SNR}_{sr_1},\ldots,\mathsf{SNR}_{sr_M},\mathsf{SNR}_{sd}+\sum_{i=1}^{M}\mathsf{SNR}_{r_id}\}$. Similar to PR, the relationship between these two effective SNRs is *indefinite*, as evidenced by the following example:

Example A.2. Consider a homogeneous network where $\mathsf{SNR}_{sr_i} = \mathsf{SNR}_{r_id} = 1$ for all $i = 1,\ldots,M$, and $\mathsf{SNR}_{sd} = \frac{1}{3}$. It follows that the effective SNR for AF is $\frac{M+1}{3}$ and the effective SNR for DF remains 1. Thus, as the number of relays increases, the effective SNR under AF scales linearly, while the effective SNR under DF is held fixed. In this case, it is clear that if $M > 2$, AF outperforms DF. Otherwise, if $M \leq 2$, AF has no advantage over DF.

APPENDIX B
PROOF OF LEMMA 1

The result in Lemma 1 was also used in [18] but without proof. Since this result is non-obvious, we provide a proof of this lemma to make this paper self-contained. Recall that $\tilde{\mathbf{H}}_{rd} \triangleq \frac{\sigma_r}{\sigma_d}\mathbf{H}_{rd}\mathbf{A}$ and $\tilde{\mathbf{H}}_{sr} \triangleq \frac{1}{\sigma_r}\mathbf{H}_{sr}\mathbf{Q}^{\frac{1}{2}}$. Thus,

we can rewrite the achievable rate expression in (6) as

$$\begin{aligned} I_{PR} &= \frac{1}{2}\log_2\left|\mathbf{I}_{N_d} + \tilde{\mathbf{H}}_{rd}\tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}(\mathbf{I}_{N_d} + \tilde{\mathbf{H}}_{rd}\tilde{\mathbf{H}}_{rd}^{\dagger})^{-1}\right| \\ &= \frac{1}{2}\log_2\left|\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}(\mathbf{I}_{N_d} + \tilde{\mathbf{H}}_{rd}\tilde{\mathbf{H}}_{rd}^{\dagger})^{-1}\tilde{\mathbf{H}}_{rd}\right| \\ &= \frac{1}{2}\log_2\left|\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger}(\mathbf{I}_{MN_r} - (\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd})^{-1})\right|, \end{aligned}$$

where the second equality follows from $|\mathbf{I}+\mathbf{MN}| = |\mathbf{I}+\mathbf{NM}|$ and the third equality follows from the fact that $\mathbf{M}(\mathbf{I}+\mathbf{NM})^{-1}\mathbf{N} = \mathbf{I} - (\mathbf{I}+\mathbf{MN})^{-1}$, and where the identities $\mathbf{I}$ are of conformable dimensions.

Using the eigenvalue decomposition of $\tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger}$ as stated in the lemma (i.e., $\tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger} = \tilde{\mathbf{U}}_{sr}\tilde{\mathbf{\Lambda}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger}$), we have

$$\begin{aligned} &\left|\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{sr}\tilde{\mathbf{H}}_{sr}^{\dagger}(\mathbf{I}_{MN_r} - (\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd})^{-1})\right| \\ &= \left|\mathbf{I}_{MN_r} + \tilde{\mathbf{U}}_{sr}\tilde{\mathbf{\Lambda}}_{sr}\tilde{\mathbf{U}}_{sr}^{\dagger}(\mathbf{I}_{MN_r} - (\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd})^{-1})\right| \\ &\overset{(a)}{=} \left|\mathbf{I}_{MN_r} + \tilde{\mathbf{\Lambda}}_{sr}\tilde{\mathbf{U}}_{sr}^{\dagger}(\mathbf{I}_{MN_r} - (\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd})^{-1})\tilde{\mathbf{U}}_{sr}\right| \\ &= \left|\mathbf{I}_{MN_r} + \tilde{\mathbf{\Lambda}}_{sr}(\mathbf{I}_{MN_r} - \tilde{\mathbf{U}}_{sr}^{\dagger}(\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd})^{-1}\tilde{\mathbf{U}}_{sr})\right| \\ &\overset{(b)}{=} \left|\mathbf{I}_{MN_r} + \tilde{\mathbf{\Lambda}}_{sr}(\mathbf{I}_{MN_r} - \tilde{\mathbf{U}}_{sr}^{-1}(\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd})^{-1}(\tilde{\mathbf{U}}_{sr}^{\dagger})^{-1})\right| \\ &= \left|\mathbf{I}_{MN_r} + \tilde{\mathbf{\Lambda}}_{sr}(\mathbf{I}_{MN_r} - (\tilde{\mathbf{U}}_{sr}^{\dagger}(\mathbf{I}_{MN_r} + \tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd})\tilde{\mathbf{U}}_{sr})^{-1})\right| \\ &= \left|\mathbf{I}_{MN_r} + \tilde{\mathbf{\Lambda}}_{sr}(\mathbf{I}_{MN_r} - (\mathbf{I}_{MN_r} + \tilde{\mathbf{U}}_{sr}^{\dagger}\tilde{\mathbf{H}}_{rd}^{\dagger}\tilde{\mathbf{H}}_{rd}\tilde{\mathbf{U}}_{sr})^{-1})\right|. \end{aligned}$$

In the above derivation, (a) again follows from $|\mathbf{I}+\mathbf{MN}| = |\mathbf{I}+\mathbf{NM}|$ and (b) exploits the fact $\tilde{\mathbf{U}}_{sr}$ is unitary so that $\tilde{\mathbf{U}}_{sr}^{\dagger} = \tilde{\mathbf{U}}_{sr}^{-1}$. Note that the last expression is exactly the same as that in the log function stated in the lemma.

APPENDIX C
DECOMPOSITIONS OF $I_{CR}(\mathbf{Q}, \mathbf{A})$

To derive the two decompositions in (24) and (25), we first need the following identities for partitioned matrices [22]:

$$\begin{vmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{U} & \mathbf{V} \end{vmatrix} = |\mathbf{X}||\mathbf{V} - \mathbf{U}\mathbf{X}^{-1}\mathbf{Y}| = |\mathbf{V}||\mathbf{X} - \mathbf{Y}\mathbf{V}^{-1}\mathbf{U}|, \tag{41}$$

where all matrices to be inverted are non-singular and all submatrices are of appropriate dimensions so that the multiplications are well-defined.

To derive the first decomposition in (24), we apply the first identity in (41) to (23). After rearranging and factoring out common terms, we obtain

$$\begin{aligned} &\left|\mathbf{I}_{2N_d} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger}\mathbf{R}^{-1}\right| = \left|\mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^{\dagger}\bar{\mathbf{R}}^{-1}\right| \left|\mathbf{I}_{N_d} + \frac{1}{\sigma_d^2} \times \right. \\ &\left. \mathbf{H}_{sd}\mathbf{Q}\left(\mathbf{I}_{N_d} - \bar{\mathbf{H}}_{sd}^{\dagger}\bar{\mathbf{R}}^{-1}\left(\mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^{\dagger}\bar{\mathbf{R}}^{-1}\right)\bar{\mathbf{H}}_{sd}\mathbf{Q}\right)\mathbf{H}_{sd}^{\dagger}\right|. \end{aligned}$$

To further simplify the above expression, we employ the celebrated Sherman-Woodbury-Morrison matrix inversion lemma: $(\mathbf{X}+\mathbf{YZ})^{-1} = \mathbf{X}^{-1} - \mathbf{X}^{-1}\mathbf{Y}(\mathbf{I}+\mathbf{ZX}^{-1}\mathbf{Y})^{-1}\mathbf{ZX}^{-1}$. By recognizing $\mathbf{X} = \mathbf{I}_{N_d}$, $\mathbf{Y} = \bar{\mathbf{H}}_{sd}^{\dagger}\bar{\mathbf{R}}^{-1}$, and

$\mathbf{Z} = \bar{\mathbf{H}}_{sd}\mathbf{Q}$ and applying the matrix inversion lemma reversely, we have

$$\begin{aligned}\left|\mathbf{I}_{2N_d} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\mathbf{R}^{-1}\right| &= \left|\mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1}\right| \times \\ &\quad \left|\mathbf{I}_{N_d} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}\mathbf{Q}\left(\mathbf{I}_{N_s} + \bar{\mathbf{H}}_{sd}^\dagger\mathbf{R}^{-1}\bar{\mathbf{H}}_{sd}\mathbf{Q}\right)^{-1}\mathbf{H}_{sd}^\dagger\right| \\ &= \left|\mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1}\right| \times \\ &\quad \left|\mathbf{I}_{N_d} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}\left(\mathbf{Q}^{-1} + \bar{\mathbf{H}}_{sd}^\dagger\mathbf{R}^{-1}\bar{\mathbf{H}}_{sd}\right)^{-1}\mathbf{H}_{sd}^\dagger\right|. \end{aligned} \tag{42}$$

Then, the decomposition in (24) follows from taking the log of both sides of (42).

To derive the decomposition in (25), we apply the second identity in (41) to (23). Similarly, by rearranging and factoring out common terms, we obtain

$$\begin{aligned}\left|\mathbf{I}_{2N_d} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\mathbf{R}^{-1}\right| &= \left|\mathbf{I}_{N_d} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}\mathbf{Q}\mathbf{H}_{sd}^\dagger\right| \left|\mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\right. \\ &\left.\left(\mathbf{I}_{N_s} - \frac{1}{\sigma_d^2}\mathbf{H}_{sd}^\dagger\left(\mathbf{I}_{N_d} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}\mathbf{Q}\mathbf{H}_{sd}^\dagger\right)^{-1}\mathbf{H}_{sd}\mathbf{Q}\right)\bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1}\right|.\end{aligned}$$

Again, by recognizing $\mathbf{X} = \mathbf{I}_{N_d}$, $\mathbf{Y} = \frac{1}{\sigma_{sd}^2}\mathbf{H}_{sd}^\dagger$, and $\mathbf{Z} = \mathbf{H}_{sd}\mathbf{Q}$ and applying the matrix inversion lemma reversely, we have

$$\begin{aligned}\left|\mathbf{I}_{2N_d} + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger\mathbf{R}^{-1}\right| &= \left|\mathbf{I}_{N_d} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}\mathbf{Q}\mathbf{H}_{sd}^\dagger\right| \times \\ &\quad \left|\mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\mathbf{Q}\left(\mathbf{I}_{N_s} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}^\dagger\mathbf{H}_{sd}\mathbf{Q}\right)^{-1}\bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1}\right| \\ &= \left|\mathbf{I}_{N_d} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}\mathbf{Q}\mathbf{H}_{sd}^\dagger\right| \times \\ &\quad \left|\mathbf{I}_{N_d} + \bar{\mathbf{H}}_{sd}\left(\mathbf{Q}^{-1} + \frac{1}{\sigma_d^2}\mathbf{H}_{sd}^\dagger\mathbf{H}_{sd}\right)^{-1}\bar{\mathbf{H}}_{sd}^\dagger\bar{\mathbf{R}}^{-1}\right|. \end{aligned} \tag{43}$$

Then, the decomposition in (25) follows from taking the log of both sides of (43).

APPENDIX D
PROOF OF LEMMA 4

Due to the almost identical structure of Problems (34) and (35), we only focus on Problem (34), i.e., deriving the maximal value of $d_j^{(i)}$. The proof for the minimal value is similar and is omitted here for brevity (in fact, the proof is identical due to exactly the same KKT systems of Problems (34) and (35)). To prove the result in Lemma 4, we first note that Problem (34) is convex and satisfies Slater's interiority condition [24]. Thus, the KKT optimality conditions are both necessary and sufficient for this problem.

Since Problem (34) is symmetric with respect to all j, without loss of generality, let us consider $\max d_1^{(i)}$ in Problem (34). For notational convenience, we omit the index i in the rest of the proof. Introducing a dual variable

$u \geq 0$ for the constraint in (34), the Lagrangian of (34) can be written as

$$L(d_1, \ldots, d_{MN_r}, u) = d_1 + u \left(P_R - \sum_{j=1}^{MN_r} (\mathbf{S})_{jj} d_j^2 - \sum_{j=2}^{MN_r} \sum_{k=1}^{j-1} 2\mathfrak{Re}\left((\mathbf{S})_{jk}\right) d_j d_k \right), \tag{44}$$

where the $\mathfrak{Re}\left(\cdot\right)$ operation means taking the real part. By KKT conditions, we have

$$\frac{\partial}{\partial d_1} L(d_1, \ldots, d_{MN_r}, u) = 1 - u \sum_{j=1}^{MN_r} 2\mathfrak{Re}\left((\mathbf{S})_{1j}\right) d_j = 0, \tag{45}$$

$$\frac{\partial}{\partial d_j} L(d_1, \ldots, d_{MN_r}, u) = -u \sum_{k=1}^{MN_r} 2\mathfrak{Re}\left((\mathbf{S})_{jk}\right) d_k = 0, \ j = 2, \ldots, MN_r. \tag{46}$$

From (45), we have $u = \frac{1}{\sum_{j=1}^{MN_r} 2\mathfrak{Re}\left((\mathbf{S})_{1j}\right) d_j} \neq 0$. Thus, from (46), we have $\sum_{k=1}^{MN_r} 2\mathfrak{Re}\left((\mathbf{S})_{jk}\right) d_k = 0$, $j = 2, \ldots, MN_r$, i.e., a homogeneous linear equation system $\bar{\mathbf{S}}^{(1)}\mathbf{d} = \mathbf{0}$, where

$$\bar{\mathbf{S}}^{(1)} \triangleq \begin{bmatrix} \mathfrak{Re}\left((\mathbf{S})_{21}\right) & \mathfrak{Re}\left((\mathbf{S})_{22}\right) & \cdots & \mathfrak{Re}\left((\mathbf{S})_{2,MN_r}\right) \\ \mathfrak{Re}\left((\mathbf{S})_{31}\right) & \mathfrak{Re}\left((\mathbf{S})_{32}\right) & \cdots & \mathfrak{Re}\left((\mathbf{S})_{3,MN_r}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathfrak{Re}\left((\mathbf{S})_{MN_r,1}\right) & \mathfrak{Re}\left((\mathbf{S})_{MN_r,2}\right) & \cdots & \mathfrak{Re}\left((\mathbf{S})_{MN_r,MN_r}\right) \end{bmatrix}, \text{ and } \mathbf{d} \triangleq \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{MN_r} \end{bmatrix}.$$

We point out that the coefficient matrix $\bar{\mathbf{S}}^{(1)}$ is obtained by taking the real parts of the entries in $\mathbf{S}$ and then deleting the first row. Thus, $\bar{\mathbf{S}}^{(1)}$ is of size $(MN_r - 1) \times MN_r$, which means that its null space is of dimension 1. Since the null space of a matrix is spanned by its right singular vectors that correspond to the zero singular values [22], we have that the solutions of $\bar{\mathbf{S}}^{(1)}\mathbf{d} = \mathbf{0}$ can be written as

$$\begin{bmatrix} d_1 & d_2 & \cdots & d_{MN_r} \end{bmatrix}^T = t\bar{\mathbf{v}}^{(1)} = t\begin{bmatrix} \bar{v}_1^{(1)} & \bar{v}_2^{(1)} & \cdots & \bar{v}_{MN_r}^{(1)} \end{bmatrix}^T, \quad t \in \mathbb{R}, \tag{47}$$

where vector $\bar{\mathbf{v}}^{(1)}$ is the right singular vector of $\bar{\mathbf{S}}^{(1)}$ corresponding to the 0 singular value. That is, in the SVD $\bar{\mathbf{S}}^{(1)} = \bar{\mathbf{U}}^{(1)}\bar{\mathbf{\Lambda}}^{(1)}(\bar{\mathbf{V}}^{(1)})^\dagger$, $\bar{\mathbf{v}}^{(1)}$ is positioned as follows:

$$\bar{\mathbf{S}}^{(1)} = \bar{\mathbf{U}}^{(1)} \begin{bmatrix} \ddots & \\ & 0 \end{bmatrix} \begin{bmatrix} \cdots & \bar{\mathbf{v}}^{(1)} \end{bmatrix}^\dagger.$$

Note that by appropriate scaling, (47) can be further equivalently written as

$$\begin{bmatrix} d_1 & d_2 & \cdots & d_{MN_r} \end{bmatrix}^T = d_1\hat{\mathbf{v}}^{(1)}, \tag{48}$$

where $\hat{\mathbf{v}}^{(1)}$ is defined as $\hat{\mathbf{v}}^{(1)} \triangleq \begin{bmatrix} 1 & \frac{\bar{v}_2^{(1)}}{\bar{v}_1^{(1)}} & \cdots & \frac{\bar{v}_{MN_r}^{(1)}}{\bar{v}_1^{(1)}} \end{bmatrix}^T$.

On the other hand, since $u \neq 0$ and by the complementary slackness condition $u(P_R - \mathbf{d}^\dagger\mathbf{S}\mathbf{d}) = 0$, we have $\mathbf{d}^\dagger\mathbf{S}\mathbf{d} = P_R$. Substituting (48) into this equality, we have $d_1^2(\hat{\mathbf{v}}^{(1)})^\dagger\mathbf{S}\hat{\mathbf{v}}^{(1)} = P_R$. It then follows that at optimality in (34), we have

$$|d_1| = \sqrt{\frac{P_R}{(\hat{\mathbf{v}}^{(1)})^\dagger\mathbf{S}\hat{\mathbf{v}}^{(1)}}},$$

which is the desired result in Lemma 4. □

REFERENCES

[1] J. N. Laneman, D. N. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity – Part I and Part II," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1948, Nov. 2003.

[3] E. C. van der Meulen, "Three-terminal communication channels," *Advanced Applied Probability*, vol. 3, pp. 120–154, 1971.

[4] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[5] E. M. Yeh and R. A. Berry, "Throughput optimal control of cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3827–3833, Oct. 2007.

[6] T. Himsoon, W. P. Siriwongpairat, Z. Han, and K. J. R. Liu, "Lifetime maximization via cooperative nodes and relay deployment in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 306–317, Feb. 2007.

[7] S. Sharma, Y. Shi, Y. T. Hou, H. D. Sherali, and S. Kompella, "Cooperative communications in multi-hop wireless networks: Joint flow routing and relay node assignment," in *Proc. IEEE INFOCOM*, San Diego, CA, Mar. 15-19, 2010, pp. 1–9.

[8] P. Liu, Z. Tao, S. Narayanan, T. Korakis, and S. S. Panwar, "CoopMAC: A cooperative MAC for wireless LANs," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 340–354, Feb. 2007.

[9] Y. Shi, S. Sharma, Y. T. Hou, and S. Kompella, "Optimal relay assignment for cooperative communications," in *Proc. ACM MobiHoc*, Hong Kong SAR, China, May 2008, pp. 3–12.

[10] J. Cai, X. Shen, J. W. Mark, and A. S. Alfa, "Semi-distributed user relaying algorithm for amplify-and-forward wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1348–1357, Apr. 2008.

[11] X. Tang and Y. Hua, "Optimal design of non-regenerative MIMO wireless relays," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1398–1407, Apr. 2007.

[12] O. Muñoz-Medina, J. Vidal, and A. Agustín, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2593–2604, Jun. 2007.

[13] W. Guan, H. Luo, and W. Chen, "Linear relaying scheme for MIMO relay system with QoS requirements," *IEEE Signal Process. Lett.*, vol. 15, pp. 697–700, 2008.

[14] A. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO relay network," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5062–5073, Oct. 2008.

[15] 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.

[16] B. Khoshnevis, W. Yu, and R. Adve, "Grassmannian beamforming for MIMO amplify-and-forward relaying," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1397–1407, Oct. 2008.

[17] Z. Fang, Y. Hua, and J. D. Koshy, "Joint source and relay optimization for non-regenerative MIMO relay," in *Proc. IEEE Workshop Sensor Array Multi-Channel Signal Processing*, Waltham, WA, Jul. 12-14, 2006, pp. 239–243.

[18] Y. Fu, L. Yang, W.-P. Zhu, and Z. Yang, "A convex optimization design of relay precoder for two-hop MIMO relay networks," in *Proc. IEEE ICC*, Cape Town, South Africa, May 23-27, 2010.

[19] Y. Rong, "Non-regenerative multicarrier MIMO relay communications based on minimization of mean-squared error," in *Proc. IEEE ICC*, Dresden, Germany, Jun. 14-18, 2009, pp. 1–5.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

[21] B. Khoshnevis, W. Yu, and R. Adve, "Grassmannian beamforming for MIMO amplify-and-forward relaying," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1397–1407, Oct. 2008.

[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge University Press, 1990.

[23] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[24] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd ed. New York, NY: John Wiley & Sons Inc., 2006.

[25] H. D. Sherali and W. P. Adams, *A Reformulation-Linearization-Technique for Solving Discrete and Continuous Nonconvex Problems*. Boston, MA: Kluwer Academic Publishing, 1999.

[26] H. D. Sherali and C. H. Tuncbilek, "A global optimization algorithm for polynomial programming problems using a reformulation-linearization technique," *Journal of Global Optimization*, vol. 2, no. 1, pp. 101–112, 1992.

[27] H. D. Sherali and H. Wang, "Global optimization of nonconvex factorable programming problems," *Mathematical Programming*, vol. 89, no. 3, pp. 459–478, 2001.