Sum Rate Maximization for Non-Regenerative MIMO Relay Networks

Cong Sun, *Member, IEEE*, Eduard Axel Jorswieck, *Senior Member, IEEE*, and Ya-xiang Yuan

***Abstract*—A multiple-antenna amplify-and-forward (AF) two-hop interference network with multiple links and multiple relays is considered. In this paper, we optimize the transmit precoders, the receiver decoders, and the relay AF matrices to maximize the achievable sum rate. First, the total signal to total interference plus noise ratio (TSTINR) maximization approach is proposed to approximate the sum rate maximization problem as a lower bound. Under individual user and individual relay transmit power constraints, an efficient alternating direction algorithm is proposed to maximize the TSTINR. Then, we modify our TSTINR model as well as the algorithm to guarantee multiple data stream transmission, by requiring the precoding matrices to have a certain number of orthogonal columns. We propose the stream selection for preprocessing, and prove that the stream selection problem to maximize the sum rate is NP-hard. Simulation results show that our proposed stream selection TSTINR model achieves much higher sum rate compared to the existing model with the same computational cost; the proposed algorithm solves the problems efficiently, and the computation time is significantly reduced.**

***Index Terms*—MIMO AF relay network, sum rate maximization, stream selection, total signal to total interference plus noise ratio.**

I. Introduction

Multiple input multiple output (MIMO) networks have been thoroughly investigated in the modern communication theory. Recently, MIMO networks assisted by relays are popular, because relays provide coverage extension, reliability enhancement and sum rate improvement [1]. The Amplify-and-Forward (AF) mode, also known as non-regenerative relaying, is simple to implement with low complexity.

Many works investigate the fundamental limits of the MIMO relay AF networks, and some of them are listed in Table I. Many studies focus on the specific networks, with a single user pair, with a single relay, or with single antenna nodes. A variety of techniques, such as Matched Filter (MF) [37], Zero Forcing (ZF) [31], [36], [37], Dirty Paper Coding (DPC) [42], Interference Alignment (IA) [28], [32] and discrete Fourier transform [31] are considered. Among them, the Minimum Mean Square Error (MMSE) technique is the most popular, where some works minimize the MSE constrained by power budgets [11], [15], [20], [38], [39], [41], some adopt the MMSE filter as the relay AF matrix or the receiver filter [24], [31], [37], and some introduce the weighted MMSE (WMMSE) model to approximate the sum rate maximization problem as they share the same KKT points [32], [43]. Various optimization methods, such as semi-definite relaxation [26], [30], [38], sequential quadratic programming [15], [27], [33], second order cone programming [5], KKT system analysis [7], [15], [20], [24], [34], [35], [38], [42], [43] and gradient method [20]–[22] are applied. Some works solve or simplify the problems by diagonolization. The analysis of relay AF matrices and precoding matrices to diagonalize the channels often works to maximize the mutual information [4], [16], [23], minimize the MSE [9], [15], [18] or optimize their lower bounds [12], [14], [24]. The alternating direction method, also known as the block coordinate descent method, is frequently used, to decompose the problem into simpler subproblems and to lower the computational cost [17], [32], [35], [38], [39], [43].

For general MIMO relay networks, [36] applies the cooperative ZF technique. Both [38] and [39] minimize the MSE, and propose different methods to solve the subproblems. Recent work [32] provides algorithms to jointly optimize users' precoders, decoders and relay AF matrices. Total leakage interference plus noise (TLIN) minimization and WMMSE models are proposed. The WMMSE model works well generally, but the computational cost is quite high. In this paper, we propose a new approximation for the sum rate maximization problem, and propose efficient algorithms which achieve higher or similar performance with significantly lower complexity.

Sometimes too many active data streams lead to a saturation of the sum rate [45]. Stream selection has been discussed in MIMO networks [45]–[51]. [46] applies the ZF receiver and selects streams independently from a set of parallel channels. For MIMO BC, [47] and [48] select streams greedily by minimizing the noise amplification and minimizing the interference, respectively. [49] studies spatially correlated MIMO networks, and selects streams by maximizing the Signal to Leakage plus Noise Ratio (SLNR). [50] treats the problem as a Degree of Freedom (DoF) feasibility problem, and selects streams while keeping them interference free. We propose a stream selection algorithm based on the new approach for MIMO relay networks, as the preprocess to maximize the sum rate.

The main contributions of the paper are listed as follows.

1) A new approach to approximate the sum rate maximization problem is proposed.

Manuscript received February 15, 2016; revised July 10, 2016; accepted August 27, 2016. Date of publication September 8, 2016; date of current version October 19, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yongming Huang. This work was supported in part by the Deutsche Forschungsgemeinschaft under Grant JO 801/13-1, in part by the National Natural Science Foundation of China under Grants 11401039, 11331012, 11471052, and in part by the fundamental research funds for the central universities under 2014RC0902.

C. Sun is with the School of Science, Beijing University of Posts and Telecommunications, Beijing, 100876, China (e-mail: suncong86@bupt.edu.cn).

E. A. Jorswieck is with the Communications Theory, Communications Laboratory, Technische Universität Dresden, Dresden, 01062, Germany (e-mail: Eduard.Jorswieck@tu-dresden.de).

Y. Yuan is with the State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China (e-mail: yyx@lsec.cc.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2607143

TABLE I
REFERENCES ABOUT MIMO RELAY AF NETWORKS

collected references		single user pair		multiple user pairs		BC/MAC[1]
		single antenna	multi-antenna	single antenna	multi-antenna	
single relay	single antenna	[2], [3]				
	multi-antenna	[4], [5]	[9]–[16]	[25]	[31]	[40]–[42]
multiple relays	single antenna	[6]–[8]	[17]	[26]–[29]		
	multi-antenna		[18]–[24]	[30]	[32]–[39]	[43]

[1]Broadcast Channel/Multiple Access Channel.

2) An efficient algorithm is developed for solving the new model, and is applied to the WMMSE model, which significantly reduces the computation time.
3) The new model and the proposed algorithm are further modified to support multiple data stream transmission. The stream selection problem to maximize the sum rate is proved to be NP-hard. Preprocessing by stream selection is introduced. The new model preprocessed by stream selection achieves much higher sum rate than the WMMSE model with the same computational cost.

Parts of the work were reported in [33]–[35] and [52]. Compared to them, the system models are more practical and better motivated. Furthermore, Section III-B3 for the novel algorithm is completely new. Finally, detailed proofs are provided for all theoretical results. Our paper is organized as follows. The system model is introduced in Section II. In Section III we set up the TSTINR maximization model to approximate the sum rate, under individual user and individual relay transmit power constraints. For this model we propose an efficient algorithm, which also applies to the WMMSE model proposed by [32]. To support multiple data streams, we develop the multiple stream TSTINR model in Section IV. Furthermore, a stream selection algorithm for preprocessing is also introduced. Simulation results are shown in Section V. It shows that the proposed new model achieves much higher sum rate than the WMMSE model with the same computational cost; the proposed new algorithm solves the WMMSE model with significantly reduced computation time.

Notation: Lowercase and uppercase boldface represent vectors and matrices, respectively. $\mathrm{Re}(a)$ and $\mathrm{Im}(a)$ mean the real part and the imaginary part of scalar a, respectively. $\mathbf{I}_d$ represents the $d \times d$ identity matrix. $\otimes$ represents the Kronecker product. $\mathrm{vec}(\mathbf{A})$ is a column vector consisting of the columns of $\mathbf{A}$. $\mathrm{Diag}\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$ represents the block diagonal matrix, where $\mathbf{A}_i$, $i = 1, \ldots, n$ are its diagonal. $\mathrm{diag}\{\mathbf{A}\}$ is a vector with elements as the diagonal elements of matrix $\mathbf{A}$. $\{0, 1\}^n$ is the n-dimensional binary vector set; $\mathbf{e}_n$ is an n-dimensional vector with all the components being 1. $\mathcal{K}$ and $\mathcal{R}$ represent the set of the user indices $\{1, 2, \ldots, K\}$ and that of the relay indices $\{1, 2, \ldots, R\}$, respectively. And we use $\mathbb{E}(\cdot)$ to denote the statistical expectation. $(\mathbf{a})_+$ means $\max(\mathbf{a}, \mathbf{0})$ componentwisely. $\nu^d_{\min}(\mathbf{A})$ is composed of the eigenvectors of $\mathbf{A}$ corresponding to its d smallest eigenvalues.

II. SYSTEM MODEL

Consider a two hop half-duplex interference channel with K user pairs and R relays, as shown in Fig. 1. Suppose Transmitter k, Receiver k and Relay r have M_k, N_k and L_r antennas, respectively, for any $k \in \mathcal{K}$ and $r \in \mathcal{R}$. User k transmits d_k parallel data streams, and the data signal vector is denoted as $\mathbf{s}_k \in \mathbb{C}^{d_k \times 1}$, where $\mathbb{E}(\mathbf{s}_k \mathbf{s}_k^H) = \mathbf{I}_{d_k}$. Here we suppose that there are no direct links between user pairs.

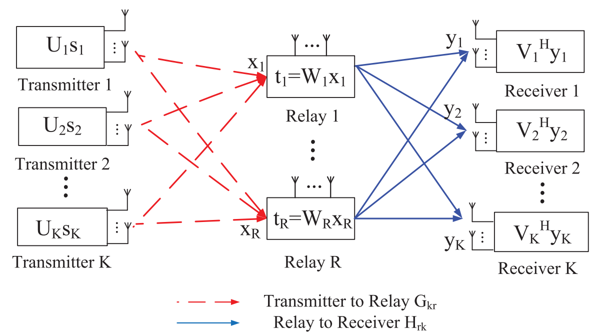


Fig. 1. MIMO relay AF network.

The communication process includes two time phases. In the first phase, each transmitter sends the precoded signal $\mathbf{U}_k \mathbf{s}_k$ to all relays, where $\mathbf{U}_k \in \mathbb{C}^{M_k \times d_k}$ is the precoding matrix of User k. Each relay receives $\mathbf{x}_r = \sum_{k \in \mathcal{K}} \mathbf{G}_{rk} \mathbf{U}_k \mathbf{s}_k + \mathbf{n}_r$. Here $\mathbf{G}_{rk} \in \mathbb{C}^{L_r \times M_k}$ is the channel matrix between Transmitter k and Relay r, and $\mathbf{n}_r$ is the noise at Relay r, with zero mean and variance matrix $\sigma_r^2 \mathbf{I}_{L_r}$. In the second phase, Relay r multiplies the received signal with its own AF matrix $\mathbf{W}_r \in \mathbb{C}^{L_r \times L_r}$, and transmits $\mathbf{t}_r = \mathbf{W}_r \mathbf{x}_r$ to all receivers. By multiplying the decoding matrix $\mathbf{V}_k \in \mathbb{C}^{N_k \times d_k}$ to the received signal, Receiver k finally obtains

$$\tilde{\mathbf{y}}_k = \underbrace{\mathbf{V}_k^H \mathbf{T}_{kk} \mathbf{s}_k}_{\text{desired signal}} + \underbrace{\sum_{q \in \mathcal{K}, q \neq k} \mathbf{V}_k^H \mathbf{T}_{kq} \mathbf{s}_q}_{\text{interference}}$$
$$+ \underbrace{\sum_{r \in \mathcal{R}} \mathbf{V}_k^H \mathbf{H}_{kr} \mathbf{W}_r \mathbf{n}_r + \mathbf{V}_k^H \mathbf{z}_k}_{\text{noise}}.$$

Here $\mathbf{H}_{kr} \in \mathbb{C}^{N_k \times L_r}$ is the channel coefficient matrix between Relay r and Receiver k, and $\mathbf{z}_k$ is the noise at Receiver k with zero mean and variance matrix $\mu_k^2 \mathbf{I}_{N_k}$. Denote the effective channel from Transmitter q to Receiver k by $\mathbf{T}_{kq} = \sum_{r \in \mathcal{R}} \mathbf{H}_{kr} \mathbf{W}_r \mathbf{G}_{rq} \mathbf{U}_q$. Suppose all the transmit signals and noises in the system are independent of each other. The transmit powers of User k and Relay r are $P_k^U = \|\mathbf{U}_k\|_F^2$, $P_r^R = \sum_{k \in \mathcal{K}} \|\mathbf{W}_r \mathbf{G}_{rk} \mathbf{U}_k\|_F^2 + \sigma_r^2 \|\mathbf{W}_r\|_F^2$, respectively.

In the following two sections we propose new efficient algorithms to design system precoders $\{\mathbf{U}\} = \{\mathbf{U}_k, k \in \mathcal{K}\}$, decoders $\{\mathbf{V}\} = \{\mathbf{V}_k, k \in \mathcal{K}\}$ and relay AF matrices $\{\mathbf{W}\} = \{\mathbf{W}_r, r \in \mathcal{R}\}$. Let $\{\mathbf{W}_{-r}\} = \{\mathbf{W}_1, \ldots, \mathbf{W}_{r-1}, \mathbf{W}_{r+1}, \ldots,$

$\mathbf{W}_R\}$ and similarly define $\{\mathbf{U}_{-k}\}$. We predefine some symbols: $\bar{\mathbf{G}}_{rk} = \mathbf{G}_{rk}\mathbf{U}_k$, $\bar{\mathbf{W}}_{rk} = \mathbf{W}_r\mathbf{G}_{rk}$, $\bar{\mathbf{H}}_{kr} = \mathbf{H}_{kr}\mathbf{W}_r$ and $\bar{\mathbf{V}}_{kr} = \mathbf{V}_k^H\mathbf{H}_{kr}$, for all $k \in \mathcal{K}$ and $r \in \mathcal{R}$.

III. Total Signal to Total Interference Plus Noise Ratio Maximization

In this section, a new TSTINR maximization model is developed to approximate the sum rate maximization problem.

A. Optimization Problem Formulation

First, we introduce a new variable, defined as

$$\text{TSTINR} = \frac{P^S}{P^I + P^N} = \frac{\sum_{k\in\mathcal{K}} P_k^S}{\sum_{k\in\mathcal{K}}(P_k^I + P_k^N)},$$

where the desired signal power, the noise power and the leakage interference at Receiver k, respectively, are expressed as $P_k^S = \mathbb{E}(\|\mathbf{V}_k^H\mathbf{T}_{kk}\mathbf{s}_k\|_F^2) = \|\mathbf{V}_k^H\sum_{r\in\mathcal{R}}\mathbf{H}_{kr}\mathbf{W}_r\mathbf{G}_{rk}\mathbf{U}_k\|_F^2$, $P_k^N = \mathbb{E}(\|\sum_{r\in\mathcal{R}}\mathbf{V}_k^H\mathbf{H}_{kr}\mathbf{W}_r\mathbf{n}_r + \mathbf{V}_k^H\mathbf{z}_k\|_F^2) = \mu_k^2\|\mathbf{V}_k\|_F^2 + \sum_{r\in\mathcal{R}}\sigma_r^2\|\mathbf{V}_k^H\mathbf{H}_{kr}\mathbf{W}_r\|_F^2$ and $P_k^I = \mathbb{E}(\|\sum_{q\in\mathcal{K},q\neq k}\mathbf{V}_k^H\mathbf{T}_{kq}\mathbf{s}_q\|_F^2) = \sum_{q\in\mathcal{K},q\neq k}\|\mathbf{V}_k^H\sum_{r\in\mathcal{R}}\mathbf{H}_{kr}\mathbf{W}_r\mathbf{G}_{rq}\mathbf{U}_q\|_F^2$. Our aim is to design the precoders $\{\mathbf{U}\}$, decoders $\{\mathbf{V}\}$ and relay AF matrices $\{\mathbf{W}\}$, in order to maximize the sum rate

$$R_{\text{sum}} = \frac{1}{2}\sum_{k\in\mathcal{K}}\log_2\det(\mathbf{I}_{N_k} + \mathbf{F}_k^{-1}\mathbf{T}_{kk}\mathbf{T}_{kk}^H) \tag{1}$$

with $\mathbf{F}_k = \sum_{q\neq k,q\in\mathcal{K}}\mathbf{T}_{kq}\mathbf{T}_{kq}^H + \sum_{r\in\mathcal{R}}\sigma_r^2\bar{\mathbf{H}}_{kr}\bar{\mathbf{H}}_{kr}^H + \mu_k^2\mathbf{I}_{N_k}$. It is complicated to optimize the sum rate directly. Here we approximate the sum rate by the newly introduced TSTINR and maximize the TSTINR instead.

If the decoding matrices $\mathbf{V}_k$, for all $k \in \mathcal{K}$, scale with a nonzero constant, the value of the TSTINR remains the same. To simplify the receive processing and to bound $\mathbf{V}_k$, we require that $\mathbf{V}_k$ has orthogonal columns. Consequently the decoding subproblems in Section III-B1 are well-defined. Furthermore, the following theorem holds:

Theorem 1: For any $\{\mathbf{U}\}$, $\{\mathbf{W}\}$ and any $\{\mathbf{V}\}$ satisfying $\mathbf{V}_k^H\mathbf{V}_k = \mathbf{I}_{d_k}, k \in \mathcal{K}$, it holds that $\frac{1}{2}\log_2[1 + \text{TSTINR}(\{\mathbf{U}\},\{\mathbf{V}\},\{\mathbf{W}\})] \leq R_{\text{sum}}(\{\mathbf{U}\},\{\mathbf{W}\})$.

The detailed proof is shown in Appendix-A. Maximizing the TSTINR provides a guaranteed system throughput. There are transmit power constraints for each user and each relay, where p_k^U and p_r^R are their power budgets, for all $k \in \mathcal{K}$ and $r \in \mathcal{R}$. The corresponding optimization problem becomes:

$$\max_{\substack{\{\mathbf{U}\},\{\mathbf{V}\},\\\{\mathbf{W}\}}} \quad \text{TSTINR} = \frac{\sum_{k\in\mathcal{K}} P_k^S}{\sum_{k\in\mathcal{K}}(P_k^I + P_k^N)} \tag{2a}$$

$$\text{s.t.} \quad \mathbf{V}_k^H\mathbf{V}_k = \mathbf{I}_{d_k}, k \in \mathcal{K}, \tag{2b}$$

$$\|\mathbf{U}_k\|_F^2 \leq p_k^U, k \in \mathcal{K}, \tag{2c}$$

$$\sum_{k\in\mathcal{K}}\|\mathbf{W}_r\mathbf{G}_{rk}\mathbf{U}_k\|_F^2 + \sigma_r^2\|\mathbf{W}_r\|_F^2 \leq p_r^R, r \in \mathcal{R}. \tag{2d}$$

Since the objective function is a fraction, there is a lack of efficient methods. Stimulated by Dinkelbach's work [56], we use a parameter C to combine the denominator and the numerator together as a new objective function[1]. Problem (2) is reformulated as:

$$\max_{\substack{\{\mathbf{U}\},\{\mathbf{V}\},\\\{\mathbf{W}\}}} \quad f(\{\mathbf{U}\},\{\mathbf{V}\},\{\mathbf{W}\};C) = C(P^I + P^N) - P^S$$

$$= \sum_{k\in\mathcal{K}}[C(P_k^I + P_k^N) - P_k^S] \tag{3a}$$

$$\text{s.t.} \quad (2b) - (2d). \tag{3b}$$

In each iteration we solve (3), and then update C as:

$$C = \frac{P^S(\{\mathbf{U}\},\{\mathbf{V}\},\{\mathbf{W}\})}{P^I(\{\mathbf{U}\},\{\mathbf{V}\},\{\mathbf{W}\}) + P^N(\{\mathbf{U}\},\{\mathbf{V}\},\{\mathbf{W}\})}. \tag{4}$$

The following theorem shows that problem (3) has a close relationship with (2) (The proof is shown in Appendix-B):

Theorem 2: If (3a) reduces in each iteration, then TSTINR is monotonically increasing. Any KKT point of (3) is a KKT point of (2).

B. Alternating Direction Algorithm

Problem (3) is a nonconvex nonlinear matrix optimization problem with many coupled variables. It is difficult to solve all the variables jointly. Here the precoders $\{\mathbf{U}\}$, the decoders $\{\mathbf{V}\}$ and the relay AF matrices $\{\mathbf{W}\}$ are solved alternatively. For each subproblem, we develop efficient methods.

1) Subproblem for the Decoding Matrix: First, we fix $\{\mathbf{U}\}$ and $\{\mathbf{W}\}$, then all the decoding matrices are independent of each other. The decoding subproblem reads:

$$\min_{\mathbf{X}\in\mathbb{C}^{N_k\times d_k}} \operatorname{tr}(\mathbf{X}^H\mathbf{A}_0\mathbf{X}) \ \text{ s.t. } \ \mathbf{X}^H\mathbf{X} = \mathbf{I}_{d_k}, \tag{5}$$

where $\mathbf{X}$ represents $\mathbf{V}_k$, and $\mathbf{A}_0 = C\mathbf{F}_k - \mathbf{T}_{kk}\mathbf{T}_{kk}^H$. The closed form solution of (5) is $\mathbf{X}^* = \nu_{\min}^{d_k}(\mathbf{A}_0)$.

2) Subproblem for the Relay AF Matrix: Given the index $r \in \mathcal{R}$, we fix $\{\mathbf{U}\}$, $\{\mathbf{V}\}$ and $\{\mathbf{W}_{-r}\}$. Thus the optimization subproblem for $\mathbf{W}_r$ is:

$$\min_{\mathbf{X}\in\mathbb{C}^{L_r\times L_r}} \quad \sum_{k\in\mathcal{K}}\operatorname{tr}\left[\mathbf{X}(\mathbf{P}_{rr}^k + C\sigma_r^2\mathbf{I}_{Lr})\mathbf{X}^H\bar{\mathbf{V}}_{kr}^H\bar{\mathbf{V}}_{kr}\right]$$

$$+ 2\text{Re}\left[\sum_{k\in\mathcal{K}}\sum_{\substack{l\neq r,\\l\in\mathcal{R}}}\operatorname{tr}(\mathbf{X}\mathbf{P}_{rl}^k\mathbf{W}_l^H\bar{\mathbf{V}}_{kl}^H\bar{\mathbf{V}}_{kr})\right]$$

$$\text{s.t.} \quad \operatorname{tr}\left[\mathbf{X}\left(\sum_{k\in\mathcal{K}}\bar{\mathbf{G}}_{rk}\bar{\mathbf{G}}_{rk}^H + \sigma_r^2\mathbf{I}_{Lr}\right)\mathbf{X}^H\right] \leq p_r^R, \tag{6}$$

where $\mathbf{P}_{rl}^k = C\sum_{q\neq k,q\in\mathcal{K}}\bar{\mathbf{G}}_{rq}\bar{\mathbf{G}}_{lq}^H - \bar{\mathbf{G}}_{rk}\bar{\mathbf{G}}_{lk}^H$, for any $k \in \mathcal{K}$ and $r, l \in \mathcal{R}$. Problem (6) is equivalent to:

$$\min_{\mathbf{p}} \ \mathbf{p}^H\bar{\mathbf{B}}_1\mathbf{p} + \bar{\mathbf{b}}^H\mathbf{p} + \mathbf{p}^H\bar{\mathbf{b}} \ \text{ s.t. } \ \mathbf{p}^H\mathbf{p} \leq p_r^R. \tag{7}$$

Here $\mathbf{p} = \mathbf{Q}\cdot\text{vec}(\mathbf{W}_r)$, $\bar{\mathbf{B}}_1 = \mathbf{Q}^{-H}\mathbf{B}_1\mathbf{Q}^{-1}$ and $\bar{\mathbf{b}} = \mathbf{Q}^{-1}\mathbf{b}$; $\mathbf{B}_1 = \sum_{k\in\mathcal{K}}(\mathbf{P}_{rr}^k + C\sigma_r^2\mathbf{I}_{L_r})^T \otimes (\bar{\mathbf{V}}_{kr}^H\bar{\mathbf{V}}_{kr})$, $\mathbf{b} = \text{vec}(\sum_{k\in\mathcal{K}}\sum_{l\neq r,l\in\mathcal{R}}\bar{\mathbf{V}}_{kr}^H\bar{\mathbf{V}}_{kl}\mathbf{W}_l(\mathbf{P}_{rl}^k)^H)$; $\mathbf{Q} \succ 0$ is computed by the

[1]The conclusions in [56] for convex sets cannot be applied to (2) directly, because its feasible set is nonconvex.

eigenvalue decomposition of $\mathbf{B}_2 = (\sum_{k\in\mathcal{K}} \bar{\mathbf{G}}_{rk}\bar{\mathbf{G}}_{rk}^H + \sigma_r^2 \mathbf{I}_{L_r})^T \otimes \mathbf{I}_L$ as $\mathbf{B}_2 = \mathbf{Q}^H\mathbf{Q}$. Let[2] $\mathbf{p}(\lambda) = -(\bar{\mathbf{B}}_1 + \lambda\mathbf{I})^{-1}\bar{\mathbf{b}}$, where $\bar{\mathbf{B}}_1 + \lambda\mathbf{I} \succeq 0$. If $\bar{\mathbf{B}}_1 \succeq 0$ and $\|\mathbf{p}(0)\|^2 \leq p_r^R$, then $\mathbf{p}(0)$ is the optimal solution of (7); otherwise, the optimal Lagrange multiplier λ is calculated by Newton's root-finding method[3] from $\|\mathbf{p}(\lambda)\|_2^2 = p_r^R$ [57, Chapter 6.1.1].

3) Subproblem for the Precoding Matrix: For User k, $\mathbf{U}_k$, we fix $\{\mathbf{V}\}$, $\{\mathbf{W}\}$ and $\{\mathbf{U}_{-k}\}$ to form the subproblem below.

$$\min_{\mathbf{X}\in\mathbb{C}^{M_k\times d_k}} \quad \mathrm{tr}(\mathbf{X}^H\mathbf{L}_k\mathbf{X})$$
$$\text{s.t.} \quad \|\mathbf{X}\|_F^2 \leq p_k^U; \mathrm{tr}(\mathbf{X}^H\bar{\mathbf{W}}_{rk}^H\bar{\mathbf{W}}_{rk}\mathbf{X}) \leq \eta_r, r\in\mathcal{R}. \quad (8)$$

Here $\mathbf{X}$ represents $\mathbf{U}_k$, $\mathbf{L}_k = \sum_{r\in\mathcal{R}}\sum_{l\in\mathcal{R}} \bar{\mathbf{W}}_{rk}^H (C\sum_{q\neq k, q\in\mathcal{K}} \bar{\mathbf{V}}_{qr}^H\bar{\mathbf{V}}_{ql} - \bar{\mathbf{V}}_{kr}^H\bar{\mathbf{V}}_{kl})\bar{\mathbf{W}}_{lk}$ and $\eta_r = p_r^R - \sigma_r^2\|\mathbf{W}_r\|_F^2 - \sum_{q\neq k, q\in\mathcal{K}}\|\bar{\mathbf{W}}_{rq}\mathbf{U}_q\|_F^2$.

a) Problem reformulation: In the following, algorithms are proposed for a more general problem. This format is applied to subproblem (8), and also the precoding subproblems of the WMMSE models in [32].

$$\min_{\mathbf{x}\in\mathbb{C}^{n\times 1}} \quad f(\mathbf{x}) = \mathbf{x}^H\mathbf{Q}_0\mathbf{x} + \mathbf{g}^H\mathbf{x} + \mathbf{x}^H\mathbf{g} \quad (9a)$$
$$\text{s.t.} \quad C_i(\mathbf{x}) = \mathbf{x}^H\mathbf{Q}_i\mathbf{x} - 1 \leq 0, i = 1, \ldots, m, \quad (9b)$$

where $\mathbf{Q}_i \succeq 0, i = 1, \ldots, m$, and $\mathbf{Q}_0$ is indefinite. Therefore (9) is a nonconvex Quadratic Constrained Quadratic Programming (QCQP). Problem (9) is equivalent to (8) with $n = M_k d_k$, $m = R+1$, $\mathbf{x} = \mathrm{vec}(\mathbf{X})$, $\mathbf{Q}_0 = \mathbf{I}_{d_k}\otimes\mathbf{L}_k$, $\mathbf{g} = \mathbf{0}$, $\mathbf{Q}_1 = \frac{1}{p_k^U}\mathbf{I}_n$ and $\mathbf{Q}_{r+1} = \frac{1}{\eta_r}\mathbf{I}_{d_k}\otimes(\bar{\mathbf{W}}_{rk}^H\bar{\mathbf{W}}_{rk})$ for all $r\in\mathcal{R}$.

To solve (9), first we propose the Feasible Shrinkage (FS) method to obtain a good interior point. Then by using it as the initialization, we apply the nonconvex Sequential Quadratic Programming (SQP) method to achieve the KKT point of (9).

Remark 1: If we apply the Semi-Definite Relaxation (SDR) method to (9), when $m > 4$ (here the relay number R is greater than 3) it is not guaranteed to recover its optimal solution [58]. The classic SQP method approximates the constraints by linearization, which is not proper for (9b). If we apply it directly and the initial point is not close to the KKT point, by experimental evidence, it leads to very short stepsizes, and thus influences the convergence and reduces the efficiency.

b) Feasible shrinkage method: The main idea of the FS method is to approximate the feasible region by an interior ellipsoid in each iteration. First, we solve the following subproblem to obtain the initial point of the FS method:

$$\min_{\mathbf{x}\in\mathbb{C}^{n\times 1}} \mathbf{x}^H\mathbf{Q}_0\mathbf{x} + \mathbf{g}^H\mathbf{x} + \mathbf{x}^H\mathbf{g} \text{ s.t. } \mathbf{x}^H\sum_{i=1}^m\mathbf{Q}_i\mathbf{x} \leq 1. \quad (10)$$

Since all $\mathbf{Q}_i$ are positive semi-definite, it holds that

$$\mathbf{x}^H\mathbf{Q}_i\mathbf{x} \leq \sum_{i=1}^m \mathbf{x}^H\mathbf{Q}_i\mathbf{x} = \mathbf{x}^H\sum_{i=1}^m\mathbf{Q}_i\mathbf{x} \leq 1, i = 1, \ldots, m. \quad (11)$$

[2]If $\bar{\mathbf{B}}_1 + \lambda\mathbf{I}$ is not invertible, $\mathbf{p}(\lambda) = -(\bar{\mathbf{B}}_1 + \lambda\mathbf{I})^{-\dagger}\bar{\mathbf{b}} + \mathbf{v}$. Here $-\dagger$ means pseudo-inverse, and $\mathbf{v}$ is a vector that belongs to the nullspace of $\bar{\mathbf{B}}_1 + \lambda\mathbf{I}$, satisfying $\|\mathbf{p}(\lambda)\|_2^2 = p_r^R$.

[3]From the computational point of view, $\frac{1}{\|\mathbf{p}(\lambda)\|_2^2} = \frac{1}{p_r^R}$ is solved [57].

Algorithm 1: Feasible Shrinkage method for (9).

input: the optimal solution of (10) as the initial point $\mathbf{x}_0$, stopping parameters ϵ_0, ϵ_1
output: $\mathbf{x}_0$
repeat
  1. Solve subproblem (12) and obtain $\mathbf{d}_0$;
  2. Update the iterative point: $\mathbf{x}_0 := \mathbf{x}_0 + \mathbf{d}_0$;
until $\|\mathbf{d}_0\|_2 < \epsilon_1$, *or there exists* $i \in \{1, 2, \ldots, m\}$, *such that* $1 - \mathbf{x}_0^T\mathbf{Q}_i\mathbf{x}_0 < \epsilon_0$;

Therefore by solving (10) we obtain a feasible point of (9), and use it as the initialization of the FS method. Next we generate the iterative step $\mathbf{d}_0$ and further update the iterative point $\mathbf{x}_0$. In each iteration, the following subproblem is solved:

$$\min_{\mathbf{d}\in\mathbb{C}^{n\times 1}} \mathbf{d}^H\mathbf{Q}_0\mathbf{d} + \bar{\mathbf{g}}^H\mathbf{d} + \mathbf{d}^H\bar{\mathbf{g}} \text{ s.t. } \mathbf{d}^H\bar{\mathbf{Q}}\mathbf{d} \leq 1, \quad (12)$$

where $\bar{\mathbf{g}} = \mathbf{g} + \mathbf{Q}_0\mathbf{x}_0$, $\bar{\mathbf{Q}} = \sum_{i=1}^m \rho_i(\mathbf{Q}_i + \rho_i\mathbf{Q}_i\mathbf{x}_0\mathbf{x}_0^H\mathbf{Q}_i)$ and $\rho_i = \frac{2}{1-\mathbf{x}_0^H\mathbf{Q}_i\mathbf{x}_0}, i = 1, \ldots, m$. Both (10) and (12) have the same structure as (7), and the same method is applied.

Theorem 3: If $\mathbf{x}_0$ is an interior feasible point of (9) and $\mathbf{d}_0$ is the optimal solution of (12), then $\mathbf{x}_0 + \mathbf{d}_0$ is also feasible for problem (9).

The detailed proof is shown in Appendix-C. Algorithm 1 presents the proposed FS method for (9).

c) Nonconvex sequential quadratic programming method: The FS method generates a good interior point for problem (9). Initialized from this point, we apply the nonconvex SQP algorithm to get a KKT point of (9).

First, we turn (9) into the real domain:

$$\min_{\hat{\mathbf{x}}\in\mathbb{R}^{2n\times 1}} \quad f(\hat{\mathbf{x}}) = \hat{\mathbf{x}}^H\hat{\mathbf{Q}}_0\hat{\mathbf{x}} + 2\hat{\mathbf{g}}^T\hat{\mathbf{x}} \quad (13a)$$
$$\text{s.t.} \quad C_i(\hat{\mathbf{x}}) = \hat{\mathbf{x}}^H\hat{\mathbf{Q}}_i\hat{\mathbf{x}} - 1 \leq 0, i = 1, \ldots, m, \quad (13b)$$

where $\hat{\mathbf{x}} = (\mathrm{Re}(\mathbf{x}^T), \mathrm{Im}(\mathbf{x}^T))^T$, $\hat{\mathbf{g}} = (\mathrm{Re}(\mathbf{g}^T), \mathrm{Im}(\mathbf{g}^T))^T$,

$$\hat{\mathbf{Q}}_0 = \begin{pmatrix} \mathrm{Re}(\mathbf{Q}_0) & -\mathrm{Im}(\mathbf{Q}_0) \\ \mathrm{Im}(\mathbf{Q}_0) & \mathrm{Re}(\mathbf{Q}_0) \end{pmatrix}, \hat{\mathbf{Q}}_i = \begin{pmatrix} \mathrm{Re}(\mathbf{Q}_i) & -\mathrm{Im}(\mathbf{Q}_i) \\ \mathrm{Im}(\mathbf{Q}_i) & \mathrm{Re}(\mathbf{Q}_i) \end{pmatrix},$$

for $i = 1, \ldots, m$. In the tth iteration, suppose the current iterative point is $\hat{\mathbf{x}}_t$. Then the following subproblem is solved:

$$\min_{\hat{\mathbf{d}}\in\mathbb{R}^{2n\times 1}} \hat{\mathbf{d}}^T\tilde{\mathbf{W}}\hat{\mathbf{d}} + 2\tilde{\mathbf{g}}^T\hat{\mathbf{d}} \text{ s.t. } \mathbf{A}^T\hat{\mathbf{d}} + \mathbf{C} \leq \mathbf{0}. \quad (14)$$

Here $\tilde{\mathbf{W}} = \hat{\mathbf{Q}}_0 + \sum_{i=1}^m \lambda_i\hat{\mathbf{Q}}_i$ is the Hessian matrix of the Lagrangian function of problem (13), where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)^T$ is the Lagrange multiplier; $\tilde{\mathbf{g}} = \hat{\mathbf{Q}}_0\hat{\mathbf{x}}_t + \hat{\mathbf{g}}$; the constraint here represents the linearization of (13b) at the point $\hat{\mathbf{x}}_t$, where the columns of $\mathbf{A}$ consist of $\hat{\mathbf{Q}}_i\hat{\mathbf{x}}_t$, for $i = 1, \ldots, m$ and $\mathbf{C} = (C_1(\hat{\mathbf{x}}_t), C_2(\hat{\mathbf{x}}_t), \ldots, C_m(\hat{\mathbf{x}}_t))^T$.

As $\tilde{\mathbf{W}}$ may be indefinite, (14) is possibly nonconvex or even unbounded.[4] If (14) is bounded, the active set method [57, Algorithm 9.4.2] is applied, and its optimal solution $\hat{\mathbf{d}}_t$

[4]This is different from the classical SQP method, where the Hessian matrices in the objective functions of the subproblems are updated by the Quasi-Newton formula and the subproblems are always convex.

Algorithm 2: Nonconvex SQP method for (13).

input: initial point $\hat{\mathbf{x}}_0$, Lagrange multiplier $\boldsymbol{\lambda} > 0$, $u = 1$, $t = 0$; stopping parameter ϵ_2
output: $\hat{\mathbf{x}}^* = \hat{\mathbf{x}}_t$
repeat
1. Solve (14) and get the iterative direction $\hat{\mathbf{d}}_t$;
2. Calculate the stepsize α_t so that $P(\hat{\mathbf{x}}_t + \alpha_t \hat{\mathbf{d}}_t, u) < P(\hat{\mathbf{x}}_t, u)$, and let $\hat{\mathbf{x}}_{t+1} := \hat{\mathbf{x}}_t + \alpha_t \hat{\mathbf{d}}_t$;
3. Update the Lagrange multiplier $\boldsymbol{\lambda}$ and the parameter u, and let $t := t + 1$;

until $\|\hat{\mathbf{d}}_t\|_2 < \epsilon_2 \|\hat{\mathbf{x}}_t\|_2$;

Algorithm 3: The hybrid algorithm to solve (9).

input: Stopping parameters ϵ_0, ϵ_1 and ϵ_2
output: $\mathbf{x}^*$
begin
1. Apply the FS method to solve (9);
2. Use the output of the FS method as the initial point, apply the nonconvex SQP method to solve (13) and get its KKT point as $\hat{\mathbf{x}}^*$;
3. Construct $\mathbf{x}^*$ from $\hat{\mathbf{x}}^*$ in the complex domain;

end

is achieved; otherwise, a direction $\hat{\mathbf{d}}_t$ is found, which satisfies $\mathbf{A}^T \hat{\mathbf{d}}_t \leq \mathbf{0}$, $\tilde{\mathbf{g}}^T \hat{\mathbf{d}}_t \leq 0$ and $\hat{\mathbf{d}}_t^T \tilde{\mathbf{W}} \hat{\mathbf{d}}_t \leq 0$. Let $P(\hat{\mathbf{x}}, u) = f(\hat{\mathbf{x}}) + u \sum_{i=1}^m \max\{C_i(\hat{\mathbf{x}}), 0\}$ be the merit function of (13), where $u > 0$ is the parameter. It is straight forward to prove that in both bounded and unbounded cases the achieved $\hat{\mathbf{d}}_t$ is a descendent direction of the merit function (See [57, Lemma 12.2.1] and [27] for the proof.). The stepsize α_t is achieved by line-search, so that $P(\hat{\mathbf{x}}_t + \alpha_t \hat{\mathbf{d}}_t, u) < P(\hat{\mathbf{x}}_t, u)$. The framework of the nonconvex SQP method is shown as Algorithm 2.

The Lagrange multiplier $\boldsymbol{\lambda}$ is first updated by the the same strategy as in [27], and then projected as $\boldsymbol{\lambda} := (\boldsymbol{\lambda})_+$. The penalty parameter u is updated by $u := \max\{u, \|\boldsymbol{\lambda}\|_\infty\}$. Since the merit function value reduces in each iteration, similar to [57, Theorem 12.2.3], we can prove that as long as the algorithm converges, the iterative point converges to a KKT point of (9). In our experiments, the nonconvex SQP always converges in several iterations (usually less than 20 iterations), because a good initialization is provided by the FS method.

d) Hybrid algorithm for the precoding subproblem: The hybrid algorithm framework to solve the equivalent subproblem (9) is summarized in Algorithm 3.

Remark 2: The constraint in subproblem (12) acts as the weighted quadratic approximation of the m constraints in (9). This forces the iterative point to go towards the boundary of the feasible region. But, in the FS method, it can never reach the boundary. Then from this point, we start the nonconvex SQP method, to achieve a KKT point of (9) locally. In our experiments, the hybrid algorithm converges very fast. See Section V for more details of the numerical illustrations.

We apply the hybrid algorithm to solve (9) and update the precoding matrices. In our TSTINR model, either the output of the FS method or the precoding matrix from the previous alternating iteration, which has lower objective function value, is used as the initial point of the nonconvex SQP method. This guarantees the reduction of (3a) in each iteration. Numerically, the initial point is always set as the output of the FS method.

Algorithm 4: Algorithm for the TSTINR model.

input: initial value of $\{\mathbf{U}\}$ and $\{\mathbf{W}\}$, $C = 1$
output: $\{\mathbf{U}\}$, $\{\mathbf{V}\}$ and $\{\mathbf{W}\}$
repeat
1. Update $\mathbf{V}_k$ by solving (5), $k \in \mathcal{K}$;
2. Update $\mathbf{W}_r$ by solving (6), $r \in \mathcal{R}$;
3. Update $\mathbf{U}_k$ by solving (8), $k \in \mathcal{K}$;
4. Update C as $C := \frac{P^S}{P^I + P^N}$;

until *Objective function value converges*;

C. Algorithm for the TSTINR Model

According to the above analysis, the framework of the algorithm to solve (2) is concluded as Algorithm 4.

The objective function reduces in each subproblem, so the value "TSTINR" converges (Theorem 2). However, as the alternating direction method is applied, there is no theoretical guarantee that the algorithm converges to a KKT point of (2).

The algorithm framework for the TSTINR is applicable to the TLIN model in [32]. (3a) is the linear combination of $P^I + P^N$ and P^S, while the TLIN model in [32] only minimizes $P^I + P^N$. From the sum rate point of view, our TSTINR model is better motivated. This is verified by simulation results, where the TSTINR always outperforms the TLIN. A similar objective function has been discussed in [59].[5] Our algorithm is also applicable to the WMMSE model [32], and saves significant computation time compared to other classical algorithms, which will be shown in Section V-A.

IV. Multiple Stream Transmission

Multiple data streams increase the data rate of single-hop networks in medium to high SNR scenarios [60]. This conclusion carries over to the two-hop case, by treating the network as an equivalent single-hop network between users and assuming the same power at relays and transmitters. In this section, we investigate the model to support the transmission of multiple data streams. We also propose an algorithm to determine the number of data streams of each user pair as a preprocessing step.

A. Analysis of the Existing Model

In simulations, we observe that the achieved precoding matrices by the TSTINR have rank one. This means only one data stream for each user pair is supported. The following theorem provides the theoretical evidence for the rank one phenomenon.[6] See Appendix-D for the proof.

Theorem 4: There always exists a rank one optimal solution for the precoding subproblem (8).

If the sum rate maximization problem is solved directly, the optimal solution should support multiple streams in high SNR

[5]In [59] the leakage interference is aimed to be aligned perfectly, thus the parameter C approaches to infinity. In our paper, $P^I + P^N$ might not reduce to zero, and consequently C might not grow to infinity.

[6]Although we cannot prove that the solution of (8) must be rank one, we always observe the rank one solution, no matter which algorithm or what initial point we use to solve (8).

scenarios. The approximation models and the used algorithms lead to the rank one precoding phenomenon. Thus in high SNR we should require the models to transmit multiple streams, in order to achieve high sum rate.

B. Multiple Stream TSTINR Model

As motivated above, the precoding matrices should have independent columns, in order to support multiple data streams. Without loss of generality all precoding matrices are required to have orthogonal columns. In our new model, we assume User k transmits with the full power p_k^U, and requires equal power allocation among d_k parallel data streams, for all $k \in \mathcal{K}$. This corresponds to the optimal power allocation scheme to maximize the sum rate in high SNR scenarios [61]. Our new optimization problem reads:

$$\max_{\substack{\{\mathbf{U}\},\{\mathbf{V}\},\\ \{\mathbf{W}\}}} \quad \text{TSTINR} = \frac{\sum_{k\in\mathcal{K}} P_k^S}{\sum_{k\in\mathcal{K}}(P_k^I + P_k^N)}$$

$$\text{s.t.} \quad \mathbf{U}_k^H \mathbf{U}_k = \frac{p_k^U}{d_k}\mathbf{I}_{d_k}, k \in \mathcal{K}; (2b), (2d). \qquad (15)$$

Similar to the model in Section III, the objective function is reformulated with the parameter C, and the alternating direction algorithm is applied. The subproblems for the decoding matrices and the relay AF matrices are the same as (5) and (6), respectively. The subproblem for the precoding matrix $\mathbf{U}_k$ has the following expression.

$$\min_{\mathbf{X}\in\mathbb{C}^{M_k \times d_k}} \quad \text{tr}(\mathbf{X}^H \mathbf{D}_0 \mathbf{X}) \qquad (16a)$$

$$\text{s.t.} \quad \mathbf{X}^H \mathbf{X} = \rho \mathbf{I}_{d_k}, \qquad (16b)$$

$$\text{tr}(\mathbf{X}^H \mathbf{D}_r \mathbf{X}) \leq \eta_r, r \in \mathcal{R}, \qquad (16c)$$

where $\mathbf{X}$ represents the matrix variable $\mathbf{U}_k$. $\rho = \frac{p_k^U}{d_k}$, $\mathbf{D}_0 = \mathbf{L}_k$ and $\mathbf{D}_r = \bar{\mathbf{W}}_{rk}^H \bar{\mathbf{W}}_{rk}$, where $\mathbf{L}_k$ and η_r are defined below (8). Because of the orthogonality constraint (16b), subproblem (16) is more complicated and more difficult to analyze than (8).

Suppose that $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_R)^T$ is the Lagrange multiplier for (16c). The dual problem of (16) is[7] [62]:

$$\min_{\boldsymbol{\mu}} h(\boldsymbol{\mu}) = \sum_{r\in\mathcal{R}} \mu_r \eta_r - \min_{\substack{\mathbf{X}\in\mathbb{C}^{M_k\times d_k},\\ \mathbf{X}^H\mathbf{X}=\rho\mathbf{I}_{d_k}}} \text{tr}\left[\mathbf{X}^H \left(\mathbf{D}_0 + \sum_{r\in\mathcal{R}} \mu_r \mathbf{D}_r\right)\mathbf{X}\right]$$

$$\text{s.t.} \quad \boldsymbol{\mu} \geq \mathbf{0}. \qquad (17)$$

Without loss of generality, we assume $\rho = 1$. The second part of $h(\boldsymbol{\mu})$ is the sum of the smallest d_k eigenvalues of $\mathbf{D}_0 + \sum_{r\in\mathcal{R}} \mu_r \mathbf{D}_r$. From [63], we derive that the subgradient of $h(\boldsymbol{\mu})$ is $\mathbf{y} = (y_1, y_2, \ldots, y_R)^T$, where $y_r = \eta_r - \text{tr}(\mathbf{X}^H \mathbf{D}_r \mathbf{X})$, for any $r \in \mathcal{R}$, and $\mathbf{X} = \nu_{\min}^{d_k}(\mathbf{D}_0 + \sum_{r\in\mathcal{R}} \mu_r \mathbf{D}_r)$, satisfying $\mathbf{X}^H\mathbf{X} = \mathbf{I}_{d_k}$. Let $\boldsymbol{\mu}^j$ and $\mathbf{y}^j$ be the iterative point and its subgradient in the jth iteration, respectively. The projected subgradient method is summarized in Algorithm IV-B, which is weakly

[7]We have converted the maximization into the minimization by multiplying the objective function with -1.

Algorithm 5: Algorithm for subproblem (16).

> **input**: Random initial point $\boldsymbol{\mu} \geq \mathbf{0}$, $\epsilon \geq 0$, $j = 1$
> **output**: the solution of (17) $\boldsymbol{\mu}^* = \boldsymbol{\mu}^j$, the solution of (16) $\mathbf{X}^* = \nu_{\min}^{d_k}(\mathbf{D}_0 + \sum_{r\in\mathcal{R}} \mu_r^* \mathbf{D}_r)$
> **repeat**
> Update the iterative point $\boldsymbol{\mu}^{j+1} = (\boldsymbol{\mu}^j - \frac{1}{j}\mathbf{y}^j)_+$ and its gradient $\mathbf{y}^{j+1}$. $j := j + 1$;
> **until** $\|\boldsymbol{\mu}^{j+1} - \boldsymbol{\mu}^j\|_2 < \epsilon$;

Algorithm 6: Algorithm for multiple stream TSTINR.

> **input**: initial value of $\{\mathbf{U}\}$ and $\{\mathbf{W}\}$, $C = 1$
> **output**: $\{\mathbf{U}\}$, $\{\mathbf{V}\}$ and $\{\mathbf{W}\}$
> **repeat**
> 1. Update $\mathbf{V}_k$ by solving (5), $k \in \mathcal{K}$;
> 2. Update $\mathbf{W}_r$ by solving (6), $r \in \mathcal{R}$;
> 3. Update $\mathbf{U}_k$ by solving (16), $k \in \mathcal{K}$;
> 4. Update C as $C := \frac{P^S}{P^I + P^N}$;
>
> **until** *Objective function value converges*;

convergent [64]. Here we stop the algorithm by measuring the distance between two consecutive iterative points, because this distance converges to zero as the iteration number j goes to infinity, due to the fact that the subgradient $\mathbf{y}^j$ is bounded.

Theorem 5: If $\boldsymbol{\mu}^*$ is the optimal solution of (17), then $\mathbf{X}^*$ is a feasible point of (16). Furthermore, if $h(\boldsymbol{\mu})$ is smooth at $\boldsymbol{\mu} = \boldsymbol{\mu}^*$, then $\mathbf{X}^*$ is the optimal solution of (16).

The detailed proof is given in Appendix-E. It shows that Algorithm IV-B achieves the optimal solution of problem (16) as long as the dual objective function is smooth at the optimal dual variable.[8] In simulations, it works well to obtain the effective precoding matrices. We summarize the algorithm to solve the multiple stream TSTINR model in Algorithm 6.

C. Stream Selection Algorithm

In MIMO relay networks, different number of data streams lead to different achievable sum rates [34]. In this subsection, we formulate the stream selection models, to maximize the sum rate by determining $\{d_k, k \in \mathcal{K}\}$. We first prove that even selecting only one stream from the simplified network is NP-hard. Then a stream selection algorithm based on the TSTINR model is proposed.

a) Computational Complexity: Consider a MIMO relay network, where each user node has single antenna. Selecting one stream is equivalent to supporting one user pair. For each user pair, we solve the following rate maximization problem with user and individual relay transmit power constraints; and we support the user transmitting with the highest rate. As $M_k = N_k = 1$ for all $k \in \mathcal{K}$, the precoding matrix $\mathbf{U}_k$ is indeed a scalar, which is denoted as u_k. Similarly, $\mathbf{H}_{kr}$ and $\mathbf{G}_{rk}$

[8]Similar to the discussion in Section IV-B, the nonsmoothness of the dual function in [35] should also be considered. Then the projected gradient method should be modified as the projected subgradient method; Theorem 1 in [35] is corrected as Theorem 5 here. The WMMSE implementation for [32] and the corresponding simulation results are also corrected.

are denoted by $\mathbf{h}_{kr}$ and $\mathbf{g}_{rk}$, respectively.

$$\max_{u_k,\{\mathbf{W}\}} \log\left(1+\frac{|\sum_{r\in\mathcal{R}}\mathbf{h}_{kr}\mathbf{W}_r\mathbf{g}_{rk}u_k|^2}{\mu_k^2+\sum_{r\in\mathcal{R}}\mathbf{h}_{kr}\mathbf{W}_r\mathbf{W}_r^H\mathbf{h}_{kr}^H}\right)$$
$$\text{s.t.}\quad |u_k|^2\le p_k^U;\|\mathbf{W}_r\mathbf{g}_{rk}u_k\|_2^2+\sigma_r^2\|\mathbf{W}_r\|_F^2\le p_r^R, r\in\mathcal{R}. \tag{18}$$

Theorem 6: Problem (18) is NP-hard.

This theorem is proved in Appendix-F. It provides the theoretical evidence that the stream selection problem to maximize the sum rate is very difficult. Furthermore, the problem is highly nonlinear and difficult to simplify. Thus we propose the stream selection algorithm based on the TSTINR model for general MIMO relay networks.[9]

b) Stream Selection Algorithm: Consider the general MIMO relay network introduced in Section II. User k could transmit at most $l_k=\min(M_k,N_k,\sum_r L_r)$ data streams. Suppose d_k out of l_k data streams are actively loaded by User k. First, we introduce $\mathbf{y}_k=(y_k^1,\ldots,y_k^{l_k})^T\in\{0,1\}^{l_k}$ to represent the status of the data streams of User k. If the ith stream is active (selected), $y_k^i=1$; otherwise $y_k^i=0$. Then $d_k=\sum_{i=1}^{l_k}y_k^i$. Let $\mathbf{Y}_k=\text{Diag}\{\mathbf{y}_k\}$. The precoding and decoding vectors of the ith data stream are the ith column of $\mathbf{U}_k\in\mathbb{C}^{M_k\times l_k}$ and that of $\mathbf{V}_k\in\mathbb{C}^{N_k\times l_k}$, respectively. For User k, the precoding and the decoding matrices consisting of the active streams are $\mathbf{U}_k\mathbf{Y}_k$ and $\mathbf{V}_k\mathbf{Y}_k$, respectively. Plugging in the new variables into the multiple stream TSTINR model, the optimization problem reads:

$$\max_{\substack{\{\mathbf{Y}\},\{\mathbf{U}\},\\ \{\mathbf{V}\},\{\mathbf{W}\}}} \frac{\tilde{P}^S}{\tilde{P}^I+\tilde{P}^N}=\frac{\sum_{k\in\mathcal{K}}\tilde{P}_k^S}{\sum_{k\in\mathcal{K}}(\tilde{P}_k^I+\tilde{P}_k^N)} \tag{19a}$$

$$\text{s.t.}\quad \mathbf{U}_k^H\mathbf{U}_k=\frac{p_k^U}{d_k}\mathbf{I}_{l_k}, k\in\mathcal{K}, \tag{19b}$$

$$\mathbf{V}_k^H\mathbf{V}_k=\mathbf{I}_{l_k}, k\in\mathcal{K}, \tag{19c}$$

$$\mathbf{Y}_k=\text{Diag}\{\mathbf{y}_k\},\mathbf{y}_k\in\{0,1\}^{l_k}, k\in\mathcal{K}, \tag{19d}$$

$$\sum_{k\in\mathcal{K}}\|\mathbf{W}_r\mathbf{G}_{rk}\mathbf{U}_k\mathbf{Y}_k\|_F^2+\sigma_r^2\|\mathbf{W}_r\|_F^2\le p_r^R, r\in\mathcal{R}. \tag{19e}$$

The expressions of $\tilde{P}_k^S$, $\tilde{P}_k^I$ and $\tilde{P}_k^N$ are the same as P_k^S, P_k^I and P_k^N in Section II, where $\mathbf{U}_k$ and $\mathbf{V}_k$ are replaced by $\mathbf{U}_k\mathbf{Y}_k$ and $\mathbf{V}_k\mathbf{Y}_k$, respectively. Here we require all the columns of $\mathbf{U}_k$ to be orthogonal, so that the precoding vectors of the active data streams are orthogonal. In (19b), $\mathbf{U}_k$ is coupled with $\mathbf{Y}_k$ through d_k. This makes the problem difficult, and the alternating direction method does not work well. To simplify the problem, we suppose that the transmit power of User k is equally allocated over all data streams, where d_k is replaced by the constant l_k in (19b). By introducing parameter C [56], the following problem is considered instead of (19).

$$\max_{\substack{\{\mathbf{Y}\},\{\mathbf{U}\},\\ \{\mathbf{V}\},\{\mathbf{W}\}}} C(\tilde{P}^I+\tilde{P}^N)-\tilde{P}^S$$
$$\text{s.t.}\quad \mathbf{U}_k^H\mathbf{U}_k=\frac{p_k^U}{l_k}\mathbf{I}_{N_k}, k\in\mathcal{K}; (19c)-(19e). \tag{20}$$

Using the alternating direction method, we fix $\{\mathbf{U}\}$, $\{\mathbf{V}\}$, $\{\mathbf{W}\}$. The subproblem on $\{\mathbf{Y}\}$ is equivalent to the following 0-1 quadratic programming problem. The detailed transformation process is shown in Appendix-G.

$$\min_{\mathbf{z}\in\{0,1\}^l} \mathbf{z}^T\mathbf{A}\mathbf{z}+\mathbf{z}^T\mathbf{a} \tag{21a}$$

$$\text{s.t.}\quad \mathbf{B}^T\mathbf{z}\le\mathbf{c}, \tag{21b}$$

Here $\mathbf{z}=(\mathbf{y}_1^T,\mathbf{y}_2^T,\ldots,\mathbf{y}_K^T)^T$ and $l_{\max}=\sum_{k\in\mathcal{K}}l_k$. $\mathbf{A}=(\tilde{\mathbf{L}}_{kq})_{K\times K}$; $\tilde{\mathbf{L}}_{kq}=C\mathbf{L}_{kq}$ for all $q\ne k,q\in\mathcal{K}$ and $\tilde{\mathbf{L}}_{kk}=-\mathbf{L}_{kk}$, for all $k\in\mathcal{K}$. Each component of $\mathbf{L}_{kq}$ is the square absolute value of the corresponding component in $\mathbf{S}_{kq}=\mathbf{V}_k^H\mathbf{T}_{kq}$. $\mathbf{a}=(\mathbf{a}_1^T,\mathbf{a}_2^T,\ldots,\mathbf{a}_K^T)^T$, where $\mathbf{a}_k=C\cdot\text{diag}(\sum_{r\in\mathcal{R}}\mathbf{D}_{kr}+\mu_k^2\mathbf{I}_{l_k})$. $\mathbf{B}$ has R columns, and the rth column $\mathbf{b}_r=(\mathbf{w}_{r1}^T,\mathbf{w}_{r2}^T,\ldots,\mathbf{w}_{rK}^T)^T$, where $\mathbf{w}_{rk}=\text{diag}\{\mathbf{U}_k^H\bar{\mathbf{W}}_{rk}^H\bar{\mathbf{W}}_{rk}\mathbf{U}_k\}$. $\mathbf{c}$ is an $R\times 1$ vector consisting of $p_r^R-\sigma_r^2\|\mathbf{W}_r\|_F^2$, $r\in\mathcal{R}$. Due to the NP-hardness, it is costly to solve (21). Furthermore, the solution of (21) relies much on the choice of $\{\mathbf{U}\},\{\mathbf{V}\},\{\mathbf{W}\}$.

Rather than solving (21) directly, we solve the following much easier subproblem.

$$\min_{\mathbf{z}=(z_1,z_2,\ldots,z_l)^T} (\mathbf{z}_0-\mathbf{z})^T\mathbf{A}(\mathbf{z}_0-\mathbf{z})+(\mathbf{z}_0-\mathbf{z})^T\mathbf{a}$$
$$\text{s.t.}\quad \mathbf{z}\in\{0,1\}^l,\sum_{i=1}^l z_i=1. \tag{22}$$

Here the dimension l first equals to $l_{\max}$, and reduces in each iteration, which is discussed in the next paragraph. $\mathbf{z}_0=\mathbf{e}_l$ represents the supported stream set, which means all the l streams are supported; $\mathbf{z}$ with only one nonzero component being 1, represents the stream to be deleted from the l streams. Since all the precoders, decoders and relay AF matrices are fixed, deleting one data stream reduces the relays' transmit power. Thus (21b) is automatically satisfied. Consequently it is omitted in (22). As (22) has only l feasible points, it is solved by exhaustive search.

Remark 3: (21) selects several streams out of the $l_{\max}$ streams; subproblem (22) deletes one stream from l streams, where first $l=l_{\max}$, and after each iteration $l:=l-1$. The main idea is that we delete streams one by one rather than selecting several streams at the same time. We first support all streams, and then delete one stream in each iteration to lower the objective function of (21).

After solving (22) and determining the stream to be deleted, we update the precoding, the decoding and the relay AF matrices accordingly: first we delete the precoding and decoding vectors of the corresponding stream; then $\{\mathbf{U}\},\{\mathbf{V}\},\{\mathbf{W}\}$ and C are updated to increase the utility function "TSTINR" by $(\{\mathbf{U}\},\{\mathbf{V}\},\{\mathbf{W}\},C):=Up(\{\mathbf{U}\},\{\mathbf{V}\},\{\mathbf{W}\})$, as one iteration of Algorithm IV-B.

In this way, we delete streams one by one, until the achieved sum rate does not increase any more. To improve the

[9] The computational complexity of the sum rate maximization problem for general MIMO relay network is clearly more interesting, and is more challenging, which is left as our future work.

Algorithm 7: Stream selection algorithm.

Input: Let $d_k = l_k$ for all $k \in \mathcal{K}$. Solve the multiple stream TSTINR model and get the initial parameters $\{\mathbf{U}^0\}, \{\mathbf{V}^0\}, \{\mathbf{W}^0\}$.
Output: $\{d\}$, $\{\mathbf{U}\}$, $\{\mathbf{V}\}$ and $\{\mathbf{W}\}$
for $j \leftarrow 1$ **to** $l_{\max}$ **do**
  Delete the jth stream in the network. $l = l_{\max} - 1$.
  $(\{\mathbf{U}\}, \{\mathbf{V}\}, \{\mathbf{W}\}, C) = Up(\{\mathbf{U}^0\}, \{\mathbf{V}^0\}, \{\mathbf{W}^0\})$;
  repeat
    1. Let $\mathbf{z}_0 = \mathbf{e}_l$. Select one stream by solving (22);
    2. Delete the corresponding precoding and decoding vector. $l = l - 1$.
    $(\{\mathbf{U}\}, \{\mathbf{V}\}, \{\mathbf{W}\}, C) = Up(\{\mathbf{U}\}, \{\mathbf{V}\}, \{\mathbf{W}\})$;
  until *the achieved sum rate is reduced*;
end
Choose the scheme with the highest sum rate.

performance, we combine the greedy idea with multiple initialization technique. In the first iteration, rather than selecting one stream to be deleted, we try every stream. This leads to l_{max} possibilities. For each possibility, we continue searching by the greedy method. Then we get l_{max} transmission schemes, and choose the best scheme with the highest achieved sum rate. This stream selection algorithm is described in Algorithm 7.

Similarly we can design the algorithm to add streams one by one, which is omitted here. Although the stream selection algorithm is based on the TSTINR model, it serves as preprocessing for any algorithm to maximize the sum rate, which provides good initial settings.

V. Simulations

In this section, we evaluate the performances of our proposed models and algorithms, where both the single stream and the multiple stream cases are considered. Each element of $\mathbf{G}_{rk}$ and $\mathbf{H}_{kr}$ obeys the complex Gaussian distribution of $\mathcal{CN}(0,1)$; the noise variances are set as $\sigma_r^2 = \sigma^2 = 1$ and $\mu_k^2 = \mu^2 = 1$, for all $r \in \mathcal{R}$ and $k \in \mathcal{K}$. Initial values of $\{\mathbf{U}\}$ and $\{\mathbf{W}\}$ are randomly generated. For the same instance we generate one random initialization for all considered models.

The parameters in the FS and the SQP methods are set to $\epsilon_0 = 10^{-6}$, $\epsilon_1 = 10^{-4}$ and $\epsilon_2 = 10^{-8}$. For each plotted point, 100 random realizations are generated to approximate the average performances. Without specific explanation, the TSTINR model runs until the achieved sum rate does not increase any more, and the WMMSE model runs until the sum rate converges. Here we define SNR as $\mathrm{SNR} = \frac{p_k^U}{\mu^2} = \frac{p_r^R}{\sigma^2}$, thus for all $r \in \mathcal{R}$ and $k \in \mathcal{K}$, $p_k^U = p_r^R = \mathrm{SNR}$. The sum rate is used as the Quality of Service (QoS) measure.

A. Single Stream Case

In this subsection, we analyze the proposed TSTINR model as well as the algorithm in Section III. All the considered networks have $d_k = 1$ to transmit single stream.

First, we consider the $(2 \times 4, 1)^4 + 2^4$ MIMO relay system.[10] For each realization, we apply our proposed TSTINR model as well as the TLIN and the WMMSE models proposed in [32]. The average achieved sum rate of the three models with respect to different SNR values is shown in Fig. 2. The TSTINR model always achieves higher sum rate than the TLIN model. For SNR greater than 25 dB, the TSTINR model achieves higher sum rate than the WMMSE model.

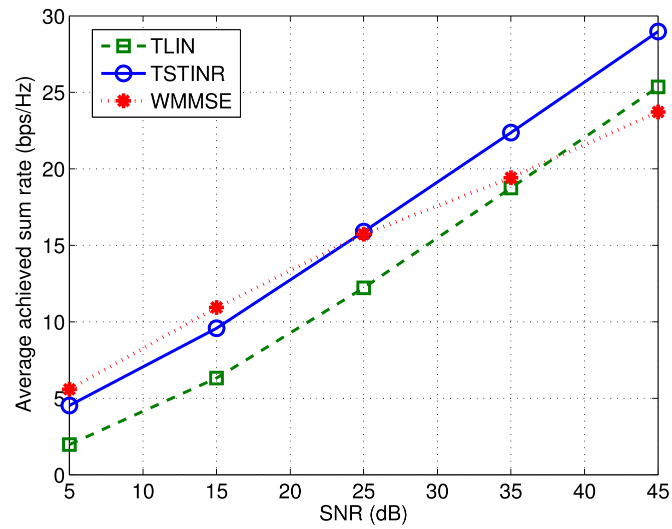


Fig. 2. Achieved sum rate for $(2 \times 4, 1)^4 + 2^4$ network.

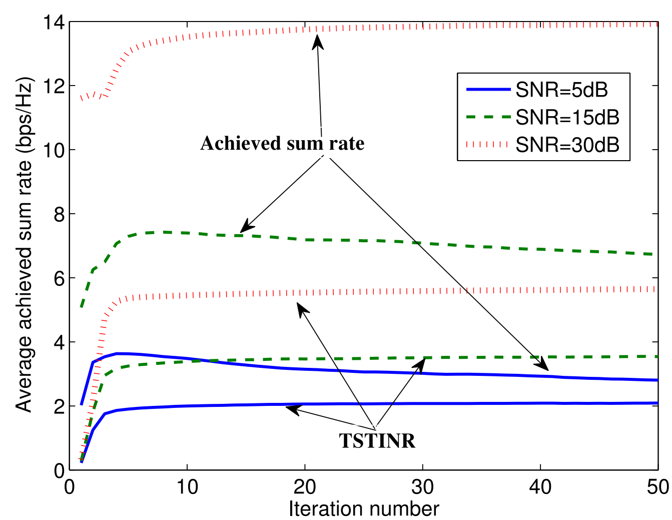


Fig. 3. Convergence of the achieved sum rate and TSTINR value.

Second, we depict the convergence curves of the TSTINR model considering the $(3 \times 3, 1)^3 + 3^3$ MIMO relay system. Fig. 3 shows the achieved sum rate over the iterations of the proposed algorithm with SNR = 5, 15, 30 dB, representing different SNR scenarios. In all the three scenarios, the achieved sum rate converges in tens of iterations. Moreover, we plot the curves representing the corresponding "TSTINR" value, to show the gap between the sum rate and the TSTINR value. For the same SNR value, the TSTINR curve is always below that of the sum rate, which validates the statement in Theorem 1.

[10]Denote $(N \times M, d)^K + L^R$ as the network with K user pairs and R relays, where each transmitter, each receiver and each relay have M, N and L antennas, respectively, and each user pair transmits d data streams. Similarly we denote $(N \times M)^K + L^R$.

TABLE II
COMPARISON OF THE PROPOSED ALGORITHM AND THE ALGORITHM IN [32]

SNR value	5dB	15dB	25dB	35dB	45dB
Achieved sum rate (bps/Hz) Algorithm [32]	8.5427	14.1383	18.1834	23.2011	28.9392
Achieved sum rate (bps/Hz) Proposed Algorithm	8.3246	14.0393	18.5745	23.5754	29.1752
Computation time (s) Algorithm [32]	62.3010	61.8395	64.8559	62.0379	62.2951
Computation time (s) (Percentage compared to Algorithm [32]) Proposed Algorithm	6.6084 (10.61%)	7.3775 (10.93%)	8.5457 (13.18%)	7.1552 (11.53%)	6.6659 (10.70%)

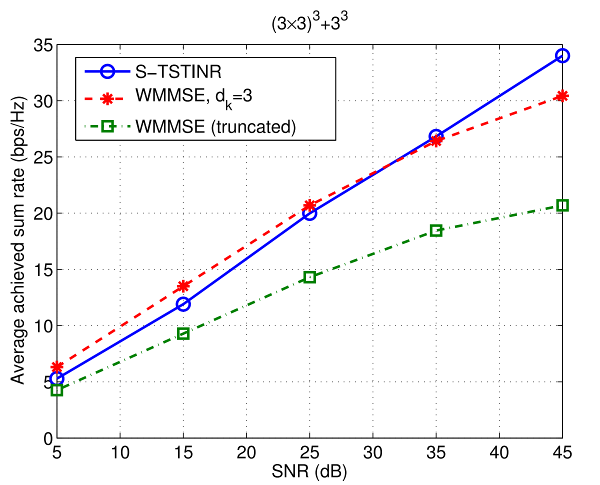


Fig. 4. Comparison between stream selected TSTINR and WMMSE: "WMMSE (truncated)," "S-TSTINR" has same computation time.

Third, we compare our proposed algorithm with other algorithms to show its efficiency. For the $(2 \times 4, 1)^4 + 4^4$ network, we solve the WMMSE model. As shown in [32], the precoding, decoding, relay AF matrices as well as the weight matrices are solved alternatively. The subproblems for the decoding matrices and the relay AF matrices are solved by the methods shown in Sections III-B1 and III-B2, respectively. The convex precoding subproblems are solved by both our proposed hybrid method Algorithm III-B3 and the software SeDuMi [66] (interior method which is suggested by [32]). The algorithms are represented by "Proposed Algorithm" and "Algorithm [32]," respectively. We stop both algorithms after 50 iterations, in order to compare their performances with constant complexity. In Table II, the achieved sum rate and the computation time of the two algorithms are listed with respect to different SNR values. Compared to "Algorithm [32]," "Proposed Algorithm" achieves almost the same sum rate, and costs less than 15% of the computation time. This shows the high efficiency of our proposed algorithm.

B. Multiple Stream Case

We investigate the multiple stream transmission models. The multiple stream TSTINR model preprocessed by the stream selection algorithm ("S-TSTINR") is compared with the WMMSE model [32] ("WMMSE"). In Fig. 4, the $(3 \times 3)^3 + 3^3$ network is considered. Here the WMMSE model is solved by our proposed algorithm from Section III-B; each user pair transmits the maximum number of data streams ($d_k = 3$, for all $k \in \mathcal{K}$), to achieve the highest sum rate. "S-TSTINR" achieves similar sum rate as "WMMSE" generally, but the computation time is much less, which is shown in Table III. We also depict the curve of "WMMSE (truncated)," representing the achieved sum rate by the WMMSE model using the same computation time as "S-TSTINR." "S-TSTINR" significantly improves the achieved sum rate, compared to "WMMSE (truncated)." For instance, at SNR $= 25$ dB, the sum rate achieved by "S-TSTINR" has 5 bps/Hz gain in sum rate. This indicates that our proposed model well balances between the QoS and the computational cost.

TABLE III
AVERAGE COMPUTATION TIME VERSUS SNR

SNR / time	5dB	15dB	25dB	35dB	45dB
Stream selected TSTINR	0.9965	1.6928	2.2419	4.1066	8.7150
WMMSE	68.9942	145.5696	201.9262	94.5086	77.4060

Remark 4: There are two reasons for the low complexity of the stream selection TSTINR model. First, the complexity of each iteration is lower, due to the projected subgradient method to update the precoders. Second, the number of iterations is usually smaller than that of the WMMSE model.

VI. CONCLUSION

This paper considers the general $K \times R \times K$ MIMO relay AF network. For the sum rate maximization problem, we propose a new approximation model called TSTINR model. With individual user and individual relay transmit power constraints, we propose an efficient algorithm to solve the TSTINR model. Next, we propose the multiple stream TSTINR model by adding the orthogonality constraints for precoders, to avoid the possible phenomenon of rank one precoding matrices. The algorithm is modified to solve the multiple stream model, where the precoding subproblem is solved by the projected subgradient method. Furthermore, the stream selection algorithm is proposed as the preprocess, and the stream selection problem to maximize the sum rate is proved to be NP-hard. The simulation results indicate that our proposed stream selection TSTINR model achieves much higher sum rate than the existing WMMSE model with the same computational cost; compared to the existing algorithm, the proposed algorithm uses less than 15% computation time, and achieves almost the same sum rate.

APPENDIX

A. Proof of Theorem 1

First we introduce two lemmas for the proof.

Lemma 1: [65, Theorem 3.2.2]

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \succeq 0, \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix} \succeq 0$$

are two Hermitian matrices with the same dimensions. $\mathbf{A}_{11} \succ 0$ and $\mathbf{B}_{11} \succ 0$ also have the same dimensions. Then

$$\frac{\det(\mathbf{A}+\mathbf{B})}{\det(\mathbf{A}_{11}+\mathbf{B}_{11})} \geq \frac{\det(\mathbf{A})}{\det(\mathbf{A}_{11})} + \frac{\det(\mathbf{B})}{\det(\mathbf{B}_{11})} \geq \frac{\det(\mathbf{B})}{\det(\mathbf{B}_{11})}. \quad (23)$$

So this derives:

$$\frac{\det(\mathbf{A}+\mathbf{B})}{\det(\mathbf{B})} \geq \frac{\det(\mathbf{A}_{11}+\mathbf{B}_{11})}{\det(\mathbf{B}_{11})}.$$

Lemma 2: [65, Theorem 6.8.1] Suppose $\mathbf{C}, \mathbf{B}$ are two Hermitian matrices. $\mathbf{C} \succeq \mathbf{B} \succ 0$. Then the following inequality holds:

$$\frac{\det(\mathbf{C})}{\det(\mathbf{B})} \geq \frac{\mathrm{tr}(\mathbf{C})}{\mathrm{tr}(\mathbf{B})}.$$

Let $\mathbf{A}_k = \mathbf{T}_{kk}\mathbf{T}_{kk}^H$, $\mathbf{B}_k = \sum_{q\in\mathcal{K}, q\neq k} \mathbf{T}_{kq}\mathbf{T}_{kq}^H + \sum_{r\in\mathcal{R}} \sigma_r^2 \bar{\mathbf{H}}_{kr}\bar{\mathbf{H}}_{kr}^H + \mu_k^2 \mathbf{I}_N$ and $\mathbf{C}_k = \mathbf{A}_k + \mathbf{B}_k, k \in \mathcal{K}$. Without loss of generality we set $\sigma_r^2 = \mu_k^2 = 1$ for all $k \in \mathcal{K}$ and $r \in \mathcal{R}$. Then

$$1 + \mathrm{TSTINR} = 1 + \frac{\sum_{k\in\mathcal{K}} \mathbf{V}_k^H \mathbf{A}_k \mathbf{V}_k}{\sum_{k\in\mathcal{K}} \mathbf{V}_k^H \mathbf{B}_k \mathbf{V}_k} = \frac{\sum_{k\in\mathcal{K}} \mathbf{V}_k^H \mathbf{C}_k \mathbf{V}_k}{\sum_{k\in\mathcal{K}} \mathbf{V}_k^H \mathbf{B}_k \mathbf{V}_k},$$
$$R_{\mathrm{sum}} = \sum_{k\in\mathcal{K}} \log_2 \det(\mathbf{B}_k^{-1}\mathbf{C}_k).$$

From the definitions we conclude that $\mathbf{B}_k \succ 0$, $\mathbf{C}_k \succ 0$ and $\mathbf{C}_k \succeq \mathbf{B}_k, k \in \mathcal{K}$.

1) We prove that for any $k \in \mathcal{K}$,

$$\det(\mathbf{B}_k^{-1}\mathbf{C}_k) \geq \det[(\mathbf{V}_k^H \mathbf{B}_k \mathbf{V}_k)^{-1}(\mathbf{V}_k^H \mathbf{C}_k \mathbf{V}_k)]. \quad (24)$$

For simplicity we omit the subscript k. Let $\mathbf{V}_\perp \in \mathbb{C}^{N\times(N-d)}$ be the bases of the complementary subspace of the subspace spanned by the columns of $\mathbf{V}$. That is, $\mathbf{Q} = [\mathbf{V}, \mathbf{V}_\perp] \in \mathbb{C}^{N\times N}$ is a unitary matrix. Then we deduce that

$$\det(\mathbf{B}^{-1}\mathbf{C}) = \frac{\det(\mathbf{Q}^H\mathbf{C}\mathbf{Q})}{\det(\mathbf{Q}^H\mathbf{B}\mathbf{Q})} \geq \frac{\det(\mathbf{V}^H\mathbf{C}\mathbf{V})}{\det(\mathbf{V}^H\mathbf{B}\mathbf{V})}.$$

The inequality is deduced from Lemma 1, $\mathbf{C} = \mathbf{A} + \mathbf{B}$ and

$$\mathbf{Q}^H\mathbf{C}\mathbf{Q} = \begin{pmatrix} \mathbf{V}^H\mathbf{C}\mathbf{V} & \mathbf{V}^H\mathbf{C}\mathbf{V}_\perp \\ \mathbf{V}_\perp^H\mathbf{C}\mathbf{V} & \mathbf{V}_\perp^H\mathbf{C}\mathbf{V}_\perp \end{pmatrix}.$$

2) $\mathbf{C}_k \succeq \mathbf{B}_k$ implies $\mathbf{V}_k^H\mathbf{C}_k\mathbf{V}_k \succeq \mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k$. Lemma 2 shows that

$$\frac{\det(\mathbf{V}_k^H\mathbf{C}_k\mathbf{V}_k)}{\det(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k)} \geq \frac{\mathrm{tr}(\mathbf{V}_k^H\mathbf{C}_k\mathbf{V}_k)}{\mathrm{tr}(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k)}. \quad (25)$$

From (24) and (25), it is concluded that

$$\begin{aligned} R_{\mathrm{sum}} &= \sum_{k\in\mathcal{K}} \log_2 \det(\mathbf{B}_k^{-1}\mathbf{C}_k) \\ &\geq \sum_{k\in\mathcal{K}} \log_2 \det[(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k)^{-1}(\mathbf{V}_k^H\mathbf{C}_k\mathbf{V}_k)] \\ &\geq \sum_{k\in\mathcal{K}} \log_2 \frac{\mathrm{tr}(\mathbf{V}_k^H\mathbf{C}_k\mathbf{V}_k)}{\mathrm{tr}(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k)}. \end{aligned} \quad (26)$$

3) Finally we prove that:

$$\sum_{k\in\mathcal{K}} \log_2 \frac{\mathrm{tr}(\mathbf{V}_k^H\mathbf{C}_k\mathbf{V}_k)}{\mathrm{tr}(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k)} \geq \log_2 \frac{\sum_{k\in\mathcal{K}} \mathrm{tr}(\mathbf{V}_k^H\mathbf{C}_k\mathbf{V}_k)}{\sum_{k\in\mathcal{K}} \mathrm{tr}(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k)}. \quad (27)$$

With any scalar $t_k \geq 1$ and the fact that $\mathrm{tr}(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k) \geq 0, k \in \mathcal{K}$, it is deduced that:

$$\left(\prod_{k\in\mathcal{K}} t_k\right) \sum_{k\in\mathcal{K}} \mathrm{tr}(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k) \geq \sum_{k\in\mathcal{K}} t_k \mathrm{tr}\left(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k\right).$$

Let $t_k = \frac{\mathrm{tr}(\mathbf{V}_k^H\mathbf{C}_k\mathbf{V}_k)}{\mathrm{tr}(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k)}, k \in \mathcal{K}$, divide $\sum_{k\in\mathcal{K}} \mathrm{tr}(\mathbf{V}_k^H\mathbf{B}_k\mathbf{V}_k)$ and take logarithm for both sides, and thus we have (27). Combining (26) and (27) we prove Theorem 1.

B. Proof of Theorem 2

Let $\{\mathbf{X}\}$ represent the set of the iterative points $\{\{\mathbf{U}\}, \{\mathbf{V}\}, \{\mathbf{W}\}\}$. Suppose $\{\mathbf{X}^i\}$ are the feasible points achieved from the ith iteration. Define $P^{I+N} = P^I + P^N$. Then the expression of parameter C used in the ith iteration as well as the TSTINR achieved in the $(i-1)$th iteration is:

$$C^i = \mathrm{TSTINR}^{i-1} = \frac{P^S(\{\mathbf{X}^{i-1}\})}{P^{I+N}(\{\mathbf{X}^{i-1}\})}.$$

Since in the ith iteration the objective function (3a) reduces, it holds that

$$\begin{aligned} f(\{\mathbf{X}^i\}; C^i) &= C^i P^{I+N}(\{\mathbf{X}^i\}) - P^S(\{\mathbf{X}^i\}) \\ &\leq f(\{\mathbf{X}^{i-1}\}; C^i) = 0, \end{aligned}$$

Then $\mathrm{TSTINR}^i = \frac{P^S(\{\mathbf{X}^i\})}{P^{I+N}(\{\mathbf{X}^i\})} \geq C^i = \mathrm{TSTINR}^{i-1}$. Thus the value of TSTINR increases monotonically.

Suppose $\{\mathbf{X}^*\} \triangleq \{\{\mathbf{U}^*\}, \{\mathbf{V}^*\}, \{\mathbf{W}^*\}\}$ are the KKT points of (3) and $\lambda_r \geq 0$ is the Lagrange multiplier of $h_r(\{\mathbf{X}\}) = (\sum_{k\in\mathcal{K}} \|\mathbf{W}_r\mathbf{G}_{rk}\mathbf{U}_k\|_F^2 + \sigma_r^2\|\mathbf{W}_r\|_F^2) - p_r^R \leq 0$, for all $r \in \mathcal{R}$. From $\{\mathbf{X}^*\}$ and the KKT conditions of (3), we shall construct the corresponding Lagrange multipliers and the KKT conditions of (2), to show that $\{\mathbf{X}^*\}$ is also a KKT point of (2). The first order optimality conditions of the problem (3) with respect to $\mathbf{W}_r$ are:

$$C\frac{\partial P^{I+N}(\{\mathbf{X}^*\})}{\partial \mathbf{W}_r} - \frac{\partial P^S(\{\mathbf{X}^*\})}{\partial \mathbf{W}_r} - \lambda_r \frac{\partial h(\{\mathbf{X}^*\})}{\partial \mathbf{W}_r} = \mathbf{0}; \quad (28)$$
$$\lambda_r h_r(\{\mathbf{X}^*\}) = 0. \quad (29)$$

Letting $C = \frac{P^S(\{\mathbf{X}^*\})}{P^{I+N}(\{\mathbf{X}^*\})}$ in (28) and dividing $-P^{I+N}$ on both sides, we obtain

$$\frac{1}{[P^{I+N}(\{\mathbf{X}^*\})]^2}\left[P^{I+N}(\{\mathbf{X}^*\})\frac{\partial P^S(\{\mathbf{X}^*\})}{\partial \mathbf{W}_r} - P^S(\{\mathbf{X}^*\})\frac{\partial P^{I+N}(\{\mathbf{X}^*\})}{\partial \mathbf{W}_r}\right] - \tilde{\lambda}_r\frac{\partial h(\{\mathbf{X}^*\})}{\partial \mathbf{W}_r} = \mathbf{0}, \quad (30)$$

where $\tilde{\lambda_r} = -\frac{1}{P^{I+N}(\{\mathbf{X}^*\})}\lambda_r$. $\tilde{\lambda}_r h_r(\{\mathbf{X}^*\}) = 0$ and (30) consist of the first order optimality conditions of (2) with respect to $\mathbf{W}_r$, for any $r \in \mathcal{R}$, where $\tilde{\lambda}_r$ is the Lagrange multiplier corresponding to the constraint (2d). Similarly, we are able to achieve the first order optimality conditions of (2) with respect to other variables. Thus $\{\mathbf{X}^*\}$ is the KKT point of (2).

C. Proof of Theorem 3

First we show Lemma 3 and prove it.

Lemma 3: $\mathbf{x} \in \mathbb{C}^{n\times 1}$ and $\mathbf{A} \succeq 0$, such that $\mathbf{x}^H\mathbf{A}\mathbf{x} < 1$, then $\mathcal{B} = \{\mathbf{y} | (\mathbf{y}-\mathbf{x})^H(\mathbf{A} + \rho\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A})(\mathbf{y}-\mathbf{x}) \leq \frac{1}{\rho}\}$ is the subset of $\mathcal{A} = \{\mathbf{y} | \mathbf{y}^H\mathbf{A}\mathbf{y} \leq 1\}$, where $\rho = \frac{2}{1-\mathbf{x}^H\mathbf{A}\mathbf{x}}$.

For any $\mathbf{y} \in \mathcal{B}$, it holds that

$$\begin{aligned}(\mathbf{y}-\mathbf{x})^H(\mathbf{A} + \rho\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A})(\mathbf{y}-\mathbf{x}) &= \mathbf{y}^H\mathbf{A}\mathbf{y} - (\mathbf{y}^H\mathbf{A}\mathbf{x} \\ &+ \mathbf{x}^H\mathbf{A}\mathbf{y})(1+\rho\mathbf{x}^H\mathbf{A}\mathbf{x}) + \rho(\mathbf{y}^H\mathbf{A}\mathbf{x})(\mathbf{x}^H\mathbf{A}\mathbf{y}) \\ &+ (\mathbf{x}^H\mathbf{A}\mathbf{x})(1+\rho\mathbf{x}^H\mathbf{A}\mathbf{x}) = \mathbf{y}^H\mathbf{A}\mathbf{y} \\ &+ \left|\sqrt{\rho}\mathbf{y}^H\mathbf{A}\mathbf{x} - \frac{1+\rho\mathbf{x}^H\mathbf{A}\mathbf{x}}{\sqrt{\rho}}\right|^2 - \frac{1}{\rho} - \mathbf{x}^H\mathbf{A}\mathbf{x} \leq \frac{1}{\rho}. \quad (31)\end{aligned}$$

From (31) and the definition of ρ, we obtain

$$\begin{aligned}\mathbf{y}^H\mathbf{A}\mathbf{y} \leq\ & \mathbf{y}^H\mathbf{A}\mathbf{y} + \left|\sqrt{\rho}\mathbf{y}^H\mathbf{A}\mathbf{x} - \frac{1+\rho\mathbf{x}^H\mathbf{A}\mathbf{x}}{\sqrt{\rho}}\right|^2 \\ \leq\ & \frac{2}{\rho} + \mathbf{x}^H\mathbf{A}\mathbf{x} = 1.\end{aligned}$$

This completes the proof of Lemma 3.

Since $\mathbf{d}_0$ is optimal for (12), the corresponding constraint holds: $\mathbf{d}_0^H\bar{\mathbf{Q}}\mathbf{d}_0 = \mathbf{d}_0^H\left[\sum_{i=1}^m \rho_i(\mathbf{Q}_i + \rho_i\mathbf{Q}_i\mathbf{x}_0\mathbf{x}_0^H\mathbf{Q}_i)\right]\mathbf{d}_0 \leq 1$. Similar to (11), for all $i = 1,\ldots,m$, it holds that $\mathbf{d}_0^H(\mathbf{Q}_i + \rho_i\mathbf{Q}_i\mathbf{x}_0\mathbf{x}_0^H\mathbf{Q}_i)\mathbf{d}_0 \leq \frac{1}{\rho_i}$. As $\mathbf{x}_0$ is an interior feasible point of (9), for any $i = 1,\ldots,m$, we have $\mathbf{x}_0^H\mathbf{Q}_i\mathbf{x}_0 < 1$. Now take $\mathbf{x}_0$, $\mathbf{x}_0 + \mathbf{d}_0$ and $\mathbf{Q}_i$ as the "$\mathbf{x}$," "$\mathbf{y}$" and "$\mathbf{A}$" in Lemma 3, respectively. We have $(\mathbf{x}_0+\mathbf{d}_0)^H\mathbf{Q}_i(\mathbf{x}_0+\mathbf{d}_0) \leq 1$, for any $i = 1,\ldots,m$, that is, $\mathbf{x}_0 + \mathbf{d}_0$ is a feasible point of (9).

D. Proof of Theorem 4

For convenience we abstract problem (8) as the following general form:

$$\min_{\mathbf{X}\in\mathbb{C}^{n\times d}} \text{tr}(\mathbf{X}^H\mathbf{A}\mathbf{X}) \text{ s.t. } \text{tr}(\mathbf{X}^H\mathbf{B}_i\mathbf{X}) \leq a_i, i = 1,2,\ldots,m. \quad (32)$$

We shall prove that, as long as $\mathbf{B}_i \succeq 0$ and $a_i > 0$, for all $i = 1,2,\ldots,m$, (32) has a rank one solution. Suppose $\mathbf{x}^*$ is the optimal solution of the following problem:

$$\min_{\mathbf{x}\in\mathbb{C}^{n\times 1}} \mathbf{x}^H\mathbf{A}\mathbf{x} \text{ s.t. } \mathbf{x}^H\mathbf{B}_i\mathbf{x} \leq a_i, i = 1,2,\ldots,m. \quad (33)$$

Let

$$\mathbf{X}^* = \frac{1}{\sqrt{d}}(\underbrace{\mathbf{x}^*, \mathbf{x}^*, \ldots, \mathbf{x}^*}_{d \text{ columns}}).$$

Obviously $\mathbf{X}^*$ has rank one. Next we prove that $\mathbf{X}^*$ is an optimal solution of problem (32) by contradiction.

If $\mathbf{X}^*$ is not optimal for (32), there must exist a feasible point $\mathbf{X}_0 = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d)$ such that $\text{tr}(\mathbf{X}_0^H\mathbf{A}\mathbf{X}_0) = \sum_{j=1}^d \mathbf{x}_j^H\mathbf{A}\mathbf{x}_j < \text{tr}[(\mathbf{X}^*)^H\mathbf{A}\mathbf{X}^*] = \sum_{j=1}^d \frac{1}{d}(\mathbf{x}^*)^H\mathbf{A}\mathbf{x}^* = (\mathbf{x}^*)^H\mathbf{A}\mathbf{x}^*$. Hence, there exists $j_0 \in \{1,2,\ldots,d\}$ such that $\mathbf{x}_{j_0}^H\mathbf{A}\mathbf{x}_{j_0} < (\mathbf{x}^*)^H\mathbf{A}\mathbf{x}^*$. Suppose $j_0 = 1$. Since $\mathbf{X}_0$ is a feasible point of (32) and $\mathbf{B}_i \succeq 0$, we can deduce that $\mathbf{x}_1^H\mathbf{B}_i\mathbf{x}_1 \leq \sum_{j=1}^d \mathbf{x}_j^H\mathbf{B}_i\mathbf{x}_j = \text{tr}(\mathbf{X}_0^H\mathbf{B}_i\mathbf{X}_0) \leq a_i$, for all $i = 1,2,\ldots,m$. Therefore $\mathbf{x}_1$ is a feasible point of problem (33), and it has smaller objective function value than $\mathbf{x}^*$. This contradicts with the fact that $\mathbf{x}^*$ is the optimal solution of (33).

So $\mathbf{X}^*$ must be an optimal solution of problem (32).

E. Proof of Theorem 5

The Lagrangian function of (17) is $L(\boldsymbol{\mu}, \mathbf{t}) = h(\boldsymbol{\mu}) - \boldsymbol{\mu}^T\mathbf{t}$, and let $\mathbf{t}^* = (t_1^*, t_2^*, \ldots, t_R^*)^T$ be the optimal Lagrange multiplier. The KKT conditions are listed below.

KKT1. $\boldsymbol{\mu} = \boldsymbol{\mu}^*$ is the first order stationary point of $\min_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, \mathbf{t}^*)$: $\eta_r - \text{tr}[(\mathbf{X}^*)^H\mathbf{D}_r\mathbf{X}^*] - t_r^* \geq 0$, for all $r \in \mathcal{R}$.

KKT2. The complementary conditions hold: for all $r \in \mathcal{R}$, $\mu_r^* t_r^* = 0$.

KKT3. The feasibility conditions hold: for all $r \in \mathcal{R}$, $\mu_r^* \geq 0$, $t_r^* \geq 0$.

The condition **KKT1** is due to the nonsmoothness of $L(\boldsymbol{\mu}, \mathbf{t})$ [57, Lemma 14.1.4]. From **KKT1** and **KKT3**, it is easy to deduce $\eta_r - \text{tr}[(\mathbf{X}^*)^H\mathbf{D}_r\mathbf{X}^*] \geq 0$, for all $r \in \mathcal{R}$. This proves that $\mathbf{X}^*$ is a feasible point of (16).

If the objective function of (17), $h(\boldsymbol{\mu})$, is smooth at $\boldsymbol{\mu} = \boldsymbol{\mu}^*$, then its subgradient becomes gradient, and the condition **KKT1** becomes:

Enhanced-KKT1. The gradient of the Lagrangian function with respect to $\boldsymbol{\mu} = \boldsymbol{\mu}^*$ is 0:

$\eta_r - \text{tr}[(\mathbf{X}^*)^H\mathbf{D}_r\mathbf{X}^*] - t_r^* = 0$, for all $r \in \mathcal{R}$.

In this case, we will prove that there is zero duality gap between (16) and (17). The duality gap is

$$\begin{aligned}&\text{tr}[(\mathbf{X}^*)^H\mathbf{D}_0\mathbf{X}^*] + h(\boldsymbol{\mu}^*) \\ &= \text{tr}[(\mathbf{X}^*)^H\mathbf{D}_0\mathbf{X}^*] + \sum_{r\in\mathcal{R}}\mu_r^*\eta_r - g(\boldsymbol{\mu}^*) \\ &= \text{tr}[(\mathbf{X}^*)^H\mathbf{D}_0\mathbf{X}^*] + \sum_{r\in\mathcal{R}}\mu_r^*\{\text{tr}[(\mathbf{X}^*)^H\mathbf{D}_r\mathbf{X}^*] + t_r^*\} \\ &\quad - \text{tr}\left[(\mathbf{X}^*)^H\left(\mathbf{D}_0 + \sum_{r\in\mathcal{R}}\mu_r^*\mathbf{D}_r\right)\mathbf{X}^*\right] = \sum_{r\in\mathcal{R}}\mu_r^* t_r^* = 0.\end{aligned}$$

The second and the last equality follow from conditions **Enhanced-KKT1** and **KKT2**, respectively. The duality gap equals 0, and consequently $\mathbf{X}^*$ is the optimal solution of (16).

F. Proof of Theorem 6

Omit the index k for simplicity, then (18) is equivalent to

$$\max_{u,\{\mathbf{W}\}} \frac{|\sum_{r\in\mathcal{R}} \mathbf{h}_r \mathbf{W}_r \mathbf{g}_r|^2 |u|^2}{\mu^2 + \sum_{r\in\mathcal{R}} \mathbf{h}_r \mathbf{W}_r \mathbf{W}_r^H \mathbf{h}_r^H}$$
$$\text{s.t.} \quad |u|^2 \le p^U; |u|^2 \|\mathbf{W}_r \mathbf{g}_r\|_2^2 + \sigma_r^2 \|\mathbf{W}_r\|_F^2 \le p_r^R, r \in \mathcal{R}. \tag{34}$$

Treat $p = |u|^2 \ge 0$ as a variable instead of u. In the following, we prove that the optimal p equals p^U, that is, the transmitter should transmit signals with the full power.

Let $\check{\mathbf{W}}_r = \mathbf{W}_r (p \cdot \mathbf{g}_r \mathbf{g}_r^H + \sigma_r^2 \mathbf{I})^{\frac{1}{2}} \sqrt{p_r^R} \sqrt{\alpha_r}$. Then problem (34) is equivalent to:

$$\max_{\substack{p;\{\check{\mathbf{W}}\};\\ \alpha_r, r\in\mathcal{R}}} \frac{p|\sum_{r\in\mathcal{R}} \mathbf{h}_r \check{\mathbf{W}}_r (p \cdot \mathbf{g}_r \mathbf{g}_r^H + \sigma_r^2 \mathbf{I})^{-\frac{1}{2}} \mathbf{g}_r (p_r^R \alpha_r)^{-\frac{1}{2}}|^2}{\mu^2 + \sum_{r\in\mathcal{R}} (p_r^R \alpha_r)^{-1} \mathbf{h}_r \check{\mathbf{W}}_r (p \cdot \mathbf{g}_r \mathbf{g}_r^H + \sigma_r^2 \mathbf{I})^{-1} \check{\mathbf{W}}_r^H \mathbf{h}_r^H}$$
$$\text{s.t.} \quad p \le p^U;\ 0 \le \alpha_r \le 1, r \in \mathcal{R}. \tag{35}$$

The numerator of the objective function is equivalent to $|\sum_{r\in\mathcal{R}} \mathbf{h}_r \check{\mathbf{W}}_r (\cdot \mathbf{g}_r \mathbf{g}_r^H + p^{-1}\sigma_r^2 \mathbf{I})^{-\frac{1}{2}} \mathbf{g}_r (p_r^R \alpha_r)^{-\frac{1}{2}}|^2$. It is easy to observe that the numerator is the monotone increasing function of p and the denominator is its monotone decreasing function. Thus the objective function is the monotone increasing function of p. The optimal p should be $p^* = p^U$. This conclusion is similar to the discussion in [7, Section III-A], where the same conclusion is analyzed for networks with single antenna relays.

By replacing $|u|^2$ with p^U, we reformulate problem (34) into the following form:

$$\max_{\mathbf{x}} \frac{\mathbf{x}^H \tilde{\mathbf{Q}} \mathbf{x}}{\mu^2 + \mathbf{x}^H \tilde{\mathbf{D}} \mathbf{x}} \ \text{s.t.}\ \mathbf{x}_r^H \mathbf{x}_r \le p_r^R, r \in \mathcal{R}. \tag{36}$$

Here $\mathbf{x} = (\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_R^T)^T$, $\mathbf{x}_r = \mathbf{J}_r^{\frac{1}{2}} \cdot \text{vec}(\mathbf{W}_r)$ and $\mathbf{w}_r = \text{vec}(\mathbf{W}_r)$; $\tilde{\mathbf{Q}} = \mathbf{JQJ}$, $\tilde{\mathbf{D}} = \mathbf{JDJ}$ and $\mathbf{J} = \text{Diag}\{\mathbf{J}_1^{-\frac{1}{2}}, \ldots, \mathbf{J}_R^{-\frac{1}{2}}\}$; $\mathbf{Q} = p^U \mathbf{t}^H \mathbf{t}$, $\mathbf{t} = ((\mathbf{g}_1 \otimes \mathbf{h}_1^T)^T, \ldots, (\mathbf{g}_R \otimes \mathbf{h}_R^T)^T)$, $\mathbf{D} = \text{Diag}(\mathbf{I}_{L_1} \otimes \mathbf{h}_1^H \mathbf{h}_1, \ldots, \mathbf{I}_{L_R} \otimes \mathbf{h}_R^H \mathbf{h}_R)$ and $\mathbf{J}_r = (p^U \mathbf{g}_r \mathbf{g}_r^H + \sigma_r^2 \mathbf{I}_{L_r})^T \otimes \mathbf{I}_{L_r}$, for all $r \in \mathcal{R}$. We derive $\mathbf{J}_r^{\frac{1}{2}} = \mathbf{U}_J \mathbf{\Lambda}_J^{\frac{1}{2}} \mathbf{U}_J^H$ from the eigenvalue decomposition of $\mathbf{J}_r \succeq 0$: $\mathbf{J}_r = \mathbf{U}_J \mathbf{\Lambda}_J \mathbf{U}_J^H$.

In the following, we shall prove that the decision problem of (36) is NP-hard. The decision problem of (36) is to check the feasibility of the following inequalities, given $\eta > 0$ [53].

$$\frac{\mathbf{x}^H \tilde{\mathbf{Q}} \mathbf{x}}{\mu^2 + \mathbf{x}^H \tilde{\mathbf{D}} \mathbf{x}} \ge \eta;\ \mathbf{x}_r^H \mathbf{x}_r \le p_r^R, r \in \mathcal{R}. \tag{37}$$

The decision problem is equivalent to solving problem (38):

$$\max_{\mathbf{x}} \mathbf{x}^H \tilde{\mathbf{B}}\, \mathbf{x} \ \text{s.t.}\ \mathbf{x}_r^H \mathbf{x}_r \le p_r^R, r \in \mathcal{R}. \tag{38}$$

Here $\tilde{\mathbf{B}} = \tilde{\mathbf{Q}} - \eta \tilde{\mathbf{D}}$. We have omitted the term $-\eta\mu^2$ in the objective function, because it is irrelevant of the variable $\mathbf{x}$.

Consider a special case of (38), where $\mathbf{x}_r = q_r \mathbf{e}_{L_r}$, and $q_r \in \mathbb{C}$ is a scalar, for all $r \in \mathcal{R}$. It becomes a quadratic programming problem with box constraints:

$$\max_{\mathbf{q}=(q_1,q_2,\ldots,q_R)^T} \mathbf{q}^H \mathbf{E} \mathbf{q} \ \text{s.t.}\ |q_r|^2 \le \frac{p_r^R}{L_r}, r \in \mathcal{R}, \tag{39}$$

where $\mathbf{E} = (e_{ij})_{R\times R}$ is an $R \times R$ matrix, $e_{ij} = \mathbf{e}_{L_i}^T \tilde{\mathbf{B}}_{ij} \mathbf{e}_{L_j}$ and $\tilde{\mathbf{B}}_{ij}$ is the ith row, jth column submatrix of the block matrix $\tilde{\mathbf{B}}$, which is partitioned in the same way as $\mathbf{x}$. Problem (39) is generally NP-hard [54]. Then the more general problem (38) is also NP-hard. From the definition of NP-hardness [53], we deduce that problem (36) and consequently problem (34) are both NP-hard.

G. Deduction of Subproblem (21)

Let $\mathbf{S}_{kq} = \mathbf{V}_k^H \mathbf{T}_{kq} \in \mathbb{C}^{l_k \times l_q}$. $\mathbf{L}_{kq}$ is constructed from $\mathbf{S}_{kq}$, where each component is the square absolute value of the corresponding component in $\mathbf{S}_{kq}$. Then we can deduce that $\|\mathbf{Y}_k^H \mathbf{V}_k^H \mathbf{T}_{kq} \mathbf{Y}_q\|_F^2 = \text{tr}(\mathbf{Y}_k^H \mathbf{S}_{kq} \mathbf{Y}_q \mathbf{Y}_q^H \mathbf{S}_{kq}^H \mathbf{Y}_k) \overset{*}{=} \text{tr}(\mathbf{Y}_k^H \mathbf{S}_{kq} \mathbf{Y}_q \mathbf{S}_{kq}^H) = \mathbf{y}_k^T \mathbf{L}_{kq} \mathbf{y}_q$. The equality with $*$ is due to the fact that $\mathbf{Y}_k = \mathbf{Y}_k^H$ and $\mathbf{Y}_k = \mathbf{Y}_k \mathbf{Y}_k^H$ for all $k \in \mathcal{K}$. Let $\tilde{\mathbf{S}}_{kr} = \sigma_r \mathbf{V}_k^H \bar{\mathbf{H}}_{kr}$, and denote $\mathbf{D}_{kr}$ as the diagonal matrix, where its diagonal elements come from diag$(\tilde{\mathbf{S}}_{kr} \tilde{\mathbf{S}}_{kr}^H)$. Then $\sigma_r^2 \|\mathbf{Y}_k^H \mathbf{V}_k^H \bar{\mathbf{H}}_{kr}\|_F^2 = \text{tr}(\mathbf{Y}_k^H \tilde{\mathbf{S}}_{kr} \tilde{\mathbf{S}}_{kr}^H \mathbf{Y}_k) = \mathbf{y}_k^H \mathbf{D}_{kr} \mathbf{y}_k = [diag(\mathbf{D}_{kr})]^T \mathbf{y}_k$.

The objective function (21a) is formulated as:

$$\begin{aligned}
&C(\tilde{P}^I + \tilde{P}^N) - \tilde{P}^S \\
&= C \sum_{k\in\mathcal{K}} \Bigg(\sum_{q\in\mathcal{K}, q\neq k} \|\mathbf{Y}_k^H \mathbf{V}_k^H \mathbf{T}_{kq} \mathbf{Y}_q\|_F^2 + \mu_k^2 \|\mathbf{V}_k \mathbf{Y}_k\|_F^2 \\
&\quad + \sum_{r\in\mathcal{R}} \sigma_r^2 \|\mathbf{Y}_k^H \mathbf{V}_k^H \bar{\mathbf{H}}_{kr}\|_F^2 \Bigg) - \sum_{k\in\mathcal{K}} \|\mathbf{Y}_k^H \mathbf{V}_k^H \mathbf{T}_{kk} \mathbf{Y}_k\|_F^2 \\
&= \sum_{k\in\mathcal{K}} \Bigg[C \sum_{q\in\mathcal{K}, q\neq k} \mathbf{y}_k^H \mathbf{L}_{kq} \mathbf{y}_q - \mathbf{y}_k^H \mathbf{L}_{kk} \mathbf{y}_k \\
&\quad + C \mathbf{y}_k^T diag\Bigg(\sum_{r\in\mathcal{R}} \mathbf{D}_{kr} + \mu_k^2 \mathbf{I}_{l_k} \Bigg) \Bigg] \\
&= \mathbf{z}^T \mathbf{A} \mathbf{z} + \mathbf{z}^T \mathbf{a},
\end{aligned}$$

where z, $\mathbf{A}$ and $\mathbf{a}$ are defined below problem (21). Similarly, the constraints (21b) are reformulated as $\mathbf{B}^T \mathbf{z} \le \mathbf{c}$.

ACKNOWLEDGMENT

The authors would like to thank the associate editor Prof. Yongming Huang and all the anonymous reviewers for their valuable suggestions. This work was partly done when Dr. Cong Sun was visiting Prof. Shiqian Ma in Chinese University of Hong Kong (CUHK), and when she was doing a postdoc in Technische Universität Dresden (TU Dresden), Germany. Dr. Cong Sun would like to thank Prof. Anthony Man-Cho So in CUHK, Prof. Yong Xia in Beihang University, China, Prof. Zi Xu in Shanghai University, China, Dr. Alessio Zappone in University of Cassino and Southern Lazio, Italy, Dr. Pin-Hsun Lin in TU Dresden, and Dr. Yafeng Liu in Chinese Academy of Sciences for their kind discussions.

References

[1] A. Høst-Madsen and A. Nosratinia, "The multiplexing gain of wireless networks," in *Proc. IEEE Int. Symp. Inf. Theory*, 2005, pp. 2065–2069.

[2] X.-D. Cai, Y.-W. Yao, and G. B. Giannakis, "Achievable rates in low-power relay links over fading channels," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 184–194, Jan. 2005.

[3] R. U. Nabar, H. Bölcskei, and F. W. Kneubühler, "Fading relay channels: Performance limits and spacetime signal design," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[4] V. Havary-Nassab, S. Shahbazpanahi, and A. Grami, "Joint receive-transmit beamforming for multi-antenna relaying schemes," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4966–4972, Sep. 2010.

[5] R. Zhang, Y.-C. Liang, C. C. Chai, and S.-G. Cui, "Optimal beamforming for two-way multi-antenna relay channel with analogue network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 699–712, Jun. 2009.

[6] G. Zheng, K.-K. Wong, A. Paulraj, and B. Ottersten, "Collaborative-relay beamforming with perfect CSI: Optimum and distributed implementation," *IEEE Signal Process. Lett.*, vol. 16, no. 4, pp. 257–260, Apr. 2009.

[7] Y. Jing and H. Hamid Jafarkhani, "Network beamforming using relays with perfect channel information," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2499–2517, Jun. 2009.

[8] L. Zhang, H.-H. Chen, and M. He, "Worst-case-based robust beamforming for wireless cooperative networks," *EURASIP J. Adv. Signal Process.*, vol. 96, pp. 1–11, 2014.

[9] 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.

[10] A. Zappone, P. Cao, and E. A. Jorswieck, "Energy efficiency optimization in relay-assisted MIMO systems with perfect and statistical CSI," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 443–457, Jan. 2014.

[11] W. Guan and H. Luo, "Joint MMSE transceiver design in non-regenerative MIMO relay systems," *IEEE Commun. Lett.*, vol. 12, no. 7, pp. 517–519, Jul. 2008.

[12] Y. Zhang, J. Li, L. Pang, and Z. Ding, "Joint optimization of source and relay precoding in non-regenerative MIMO relay systems," *Wireless Commun. Mobile Comput.*, vol. 2013, no. 13, pp. 1571–1585, 2013.

[13] J. Wagner, J. Rankov, and A. Wittneben, "Large n analysis of amplify-and-forward MIMO relay channels with correlated rayleigh fading," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5735–5746, Dec. 2008.

[14] 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.

[15] R. Mo and Y. H. Chew, "MMSE-based joint source and relay precoding design for amplify-and-forward MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4668–4676, Sep. 2009.

[16] 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.

[17] C. Li, X. Wang, L. Yang, and W.-P. Zhu, "Joint source-and-relay beamforming for multiple-input multiple-output systems with single-antenna distributed relays," *IET Commun.*, vol. 6, no. 1, pp. 62–68, Feb. 2012.

[18] L. Sanguinetti, A. A. D'Amico, and Y. Rong, "A tutorial on the optimization of amplify-and-forward MIMO relay systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1331–1346, Sep. 2012.

[19] H. Bölcskei, R. U. Nabar, Ö. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[20] C. Zhao and B. Champagne, "Joint design of multiple non-regenerative MIMO relaying matrices with power constraints," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4861–4873, Oct. 2013.

[21] K.-J. Lee, H. Sung, E. Park, and I. Lee, "Joint optimization for one and two-way MIMO AF multiple-relay systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3671–3681, Dec. 2010.

[22] A. Toding, M. K. Khandaker, and Y. Rong, "Joint source and relay design for MIMO multi-relay systems using projected gradient approach," *EURASIP J. Wireless Commun. Netw.*, vol. 151, pp. 1–9, 2014.

[23] J. Liu, N. B. Shroff, and H. D. Sherali, "Optimal power allocation in multi-relay MIMO cooperative networks: Theory and algorithms," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 331–340, Feb. 2012.

[24] Y.-H. Fu, L.-X. Yang, W.-P. Zhu, and C. Liu, "Optimum linear design of two-hop MIMO relay networks with QoS requirements," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2257–2269, May 2011.

[25] S. Berger, M. Kuhn, A. Wittneben, T. Unger, and A. Klein, "Recent advances in amplify-and-forward two-hop relaying," *IEEE Commun. Mag.*, vol. 47, no. 7, pp. 50–56, Jul. 2009.

[26] S. Fazeli-Dehkordy, S. Shahbazpanahi, and S. Gazor, "Multiple peer-to-peer communications using a network of relays," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3053–3062, Aug. 2009.

[27] C. Sun and Y. Yuan, "A fast algorithm for beamforming problems in distributed communication of relay networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 3304–3307.

[28] H. Ning, C. Ling, and K. K. Leung, "Relay-aided interference alignment: Feasibility conditions and algorithm," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 390–394.

[29] T. Gou, S. A. Jafar, C. Wang, S.-W. Jeon, and S.-Y. Chung, "Aligned interference neutralization and the degrees of freedom of the $2 \times 2 \times 2$ interference channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4381–4395, Jul. 2012.

[30] B. K. Chalise and L. Vandendorpe, "Optimization of MIMO relays for multipoint-to-multipoint communications: Nonrobust and robust designs," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6355–6368, Dec. 2010.

[31] J. Li, F. Roemer, and M. Haardt, "Spectrum and infrastructure sharing in the MIMO interference channels," in *Proc. 19th Eur. Signal Process. Conf.*, Barcelona, Spain, 2011, pp. 181–185.

[32] K. T. Truong, P. Sartori, and R. W. Heath, Jr., "Cooperative algorithms for MIMO amplify-and-forward relay networks," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1272–1287, Mar. 2013.

[33] C. Sun and E. A. Jorswieck, "Low complexity high throughput algorithms for MIMO AF relay networks," in *Proc. IEEE Int. Conf. Commun.*, Budapest, Hungary, Jun. 2013, pp. 5511–5516.

[34] C. Sun and E. A. Jorswieck, "Multi-stream sum rate maximization for MIMO AF relay networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, BC, Canada, 2013, pp. 4434–4438.

[35] C. Sun and E. A. Jorswieck, "Sum rate maximization model of non-regenerative multi-stream multi-pair multi-relay network," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brisbane, QLD, Australia, Apr. 2015, pp. 3038–3042.

[36] R. S. Ganesan, H. Al-Shatri, T. Weber, and A. Klein, "Cooperative zero forcing in multi-pair multi-relay networks," in *Proc. Int. Symp. Pers., Indoor Mobile Commun. Syst.*, Sydney, NSW, Australia, Sep. 2012, pp. 1740–1745.

[37] O. Oyman and A. Paulraj, "Design and analysis of linear distributed MIMO relaying algorithms," *Proc. Inst. Electr. Eng.–Commun.*, vol. 153, no. 4, pp. 565–572, Aug. 2006.

[38] C. Xing, S. Li, Z. Fei, and J. Kuang, "How to understand linear minimum mean-square-error transceiver design for multiple-input-multiple-output systems from quadratic matrix programming," *IET Commun.*, vol. 7, no. 12, pp. 1231–1242, Aug. 2013.

[39] K. X. Nguyen, Y. Rong, and S. Nordholm, "MMSE-based joint source and relay optimization for interference MIMO relay systems," *EURASIP J. Wireless Commun. Netw.*, vol. 73, pp. 1–9, 2015.

[40] X.-N. Wang, X.-X. Wang, W.-R Gong, and Z. Huang, "Generalized block-diagonalization schemes for MIMO relay broadcasting systems," *Int. J. Digit. Multimedia Broadcast.*, vol. 2014, 2014, Art. no. 305846, pages 6, doi: 10.1155/2014/305846

[41] H. Stefanovic, V. Stankovic, M. C. Stefanovic, P. Spalevic, S. P. Panic, and S. Milosavljevic, "Multi-user MIMO MMSE non-regenerative relaying using local channel state information," *EURASIP J. Adv. Signal Process.*, vol. 186, pp. 1–8, Jul. 2012.

[42] Y. Yu and Y. Hua, "Power allocation for a MIMO relay system with multiple-antenna users," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2823–2835, May 2010.

[43] H-J. Choi, K-J. Lee, C. Song, H. Song, and I. Lee, "Weighted sum rate maximization for multi-user multi-relay MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 885–889, Feb. 2013.

[44] C. Xing, S. Ma, Z. Fei, Y.-C. Wu, and H. V. Poor, "A general robust linear transceiver design for amplify-and-forward multi-hop MIMO relaying systems," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1196–1209, Mar. 2013.

[45] R. Hunger, D. A. Schmidt, and M. Joham, "A combinatorial approach to maximizing the sum rate in the MIMO BC with linear precoding," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Oct. 2008, pp. 316–320.

[46] R. W. Heath, M. Airy, and A. J. Paulraj, "Multiuser diversity for MIMO wireless systems with linear receivers," in *Proc. IEEE Asilomar Conf.*, Nov. 2001, pp. 1194–1199.

[47] M. Amara, D. Slock, and Y. Wu, "Recursive stream selection for CF MU-MIMO BC precoder design," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint-Petersburg, Russia, Jul. 2011, pp. 2979–2983.

[48] H. An, M. Mohaisen, D. Han, and K. Chang, "Coordinated transmit and receive processing with adaptive multi-stream selection," in *Proc. IEEE Int. Conf. Commun., Netw. Mobile Comput.*, Chengdu, China, 2010, pp. 1–5.

[49] J.-S. Kim, K.-J. Lee, H. Park, and I. Lee, "Transmission mode selection algorithms for spatially correlated MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4475–4479, Aug. 2012.

[50] R. Srinivasan, D. M. Blough, L. M. Cortés-Peña, and P. Santi, "Maximizing throughput in MIMO networks with variable rate streams," in *Proc. Eur. Wireless Conf.*, Lucca, Italy, Apr. 2010, pp. 551–559.

[51] S.-Q. Zhang and V. K. N. Lau, "Multi-relay selection design and analysis for multi-stream cooperative communications," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1082–1089, Apr. 2011.

[52] C. Sun and E. A. Jorswieck, "Stream selection methods for MIMO relay AF networks," in *Proc. IEEE Conf. Sensor Array Multichannel Signal Process. Workshop*, Rio, Jul. 2016, pp. 1–5.

[53] D.-Z. Du and K.-I. Ko, *Theory of Computational Complexity*. Hoboken, NJ, USA: Wiley, 2000.

[54] S. Burer and A. N. Letchford, "On non-convex quadratic programming with box constraints," *SIAM J. Optim.*, vol. 20, no. 2, pp. 1073–1089, 2009.

[55] C. S. Vaze and M. K. Varanasi, "Beamforming and aligned interference neutralization achieve the degrees of freedom region of the $2 \times 2 \times 2$ MIMO interference network," in *Proc. Inf. Theory Appl. Workshop*, CA, USA, 2012, pp. 199–203.

[56] W. Dinkelbach, "On nonlinear fractional programming," *Manag. Sci.*, vol. 13, no. 7, pp. 492–498, 1967.

[57] W.-Y. Sun and Y.-X. Yuan, *Optimization Theory and Methods: Nonlinear Programming.* New York, NY, USA: Springer-Verlag, 2006, pp. 130–135, 303–308.

[58] W. Ai, Y. Huang, and S. Zhang, "New results on Hermitian matrix rank-one decomposition," *Math. Program.*, vol. 128, pp. 253–283, 2011.

[59] C. Sun, Y. Yang, and Y. Yuan, "Low complexity interference alignment algorithms for desired signal power maximization problem of MIMO channels," *EURASIP J. Adv. Signal Process.*, vol. 137, Jul. 2012, pp. 1–13, doi: 10.1186/1687-6180-2012-137.

[60] W. Yu and W. Rhee, "Degrees of freedom in wireless multiuser spatial multiplex systems with multiple antennas," *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1747–1753, Oct. 2006.

[61] W. Rhee and J.-M. Cioffi, "On the capacity of multiuser wireless channels with multiple antennas," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2580–2595, Oct. 2003.

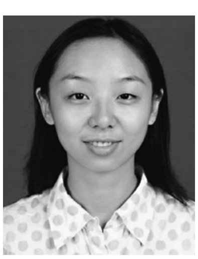
[62] S. Boyd and L. Vandenberghe, *Convex Optimization.* New York, NY, USA: Cambridge Univ. Press, 2004, pp. 215–264.

[63] C. S. Rudisill, "Derivatives of eigenvalues and eigenvectors for a general matrix," *AIAA J.*, vol. 12, no. 5, pp. 721–722, 1974.

[64] Y. I. Albor, A. N. Iusem, and M. V. Solodov, "On the projected subgradient method for nonsmooth convex optimization in a Hilbert space," *Math. Program.*, vol. 81, no. 1, pp. 23–35, Mar. 1998.

[65] S. Wang, M. Wu, and Z. Jia, *Matrix Inequalities*, 2nd ed. (in Chinese). Beijing, China: Science Press, 2006, pp. 54–57, 157–158.

[66] J. F. Sturm, "Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, pp. 625–653, 1999.



Cong Sun (M'13) received the B.S. degree in applied mathematics from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2008 and the Ph.D. degree in operations research from Chinese Academy of Sciences, Beijing, in 2013. She visited TU Dresden in 2012 and in 2015. She has been a Lecturer with the School of Science, BUPT, since 2013. Her research interests include MIMO system, multiple hop network, interference alignment, and nonlinear optimization techniques.

Eduard Axel Jorswieck (S'01–M'03–SM'08) received the Diplom-Ingenieur (M.S.) degree and Doktor-Ingenieur (Ph.D.) degree, both in electrical engineering and computer science, from the Technische Universität Berlin (TU Berlin), Berlin, Germany, in 2000 and 2004, respectively.

He was with the Broadband Mobile Communication Networks Department, Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, Berlin, from 2000 to 2008. From 2005 to 2008, he was a Lecturer with the TU Berlin. From 2006 to 2008, he was with the Department of Signals, Sensors and Systems, Royal Institute of Technology, as a Postdoctoral Researcher and an Assistant Professor. Since 2008, he has been the Head of the Chair of Communications Theory and a Full Professor with the Technische Universität Dresden, Germany. He is a principal Investigator in the Excellence Cluster Center for Advancing Electronics Dresden and the founding Member of the 5G lab Germany. His main research interests include the area of signal processing for communications and networks, applied information theory, and communications theory. He has authored more than 85 journal papers, 10 book chapters, some 230 conference papers, and 3 monographs on these research topics.

Dr. Jorswieck received the IEEE Signal Processing Society Best Paper Award in 2006 and coauthored papers that won the Best Paper or Best Student Paper Awards at IEEE WPMC 2002, Chinacom 2010, IEEE CAMSAP 2011, IEEE SPAWC 2012, and IEEE WCSP 2012. He was a Member of the IEEE SPCOM Technical Committee during 2008–2013, and has been a Member of the IEEE SAM Technical Committee since 2015. He has served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2011 to 2015. Since 2008, continuing until 2011, he has served as an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS, and until 2013, as a Senior Associate Editor. Since 2013, he has been an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Ya-xiang Yuan received the B.Sc. degree in computational mathematics in 1981 from Xiangtan University, Hunan, China, and the Ph.D. degree in applied mathematics in 1986 from the University of Cambridge, Cambridge, U.K.

From 1985 to 1988, he was a Retherford Research Fellow at Fitzwilliam College, University of Cambridge. In 1988, he became a Full Professor in the Computing Center of the Chinese Academy of Sciences (CAS). He was the Director of the Institute of Computational Mathematics of CAS during 1995–2006, the Director of the State Key Laboratory of Scientific and Engineering Computing during 1997–2005, and the Vice President of the Academy of Mathematics and Systems Science of CAS during 1998–2006. His main research interest includes nonlinear optimization theory, including trust region method, subspace method, quasi-Newton method, etc. He received numerous awards, including Fox Prize (1985, London), Young Scientist Award of China (1996, Beijing), Second Prize of National Natural Science Award (2006, Beijing), S.S. Chern Award of Chinese Mathematical Society (2011), the TWAS Prize in mathematics (2014), and the SU Buqing Prize of China SIAM (2016). He gave a plenary lecture at the ICIAM 1999, and an invited lecture at the ICM 2014. He was elected as an academician of CAS in 2011. He became a Fellow of SIAM in 2011, a Fellow of AMS in 2012, a Corresponding Member of Brazilian Academy of Science in 2014, and a Fellow of the TWAS in 2015. He is currently the President of Chinese Mathematical Society, the Editor-in-Chief of "*Science China: Mathematics,*" and the Editor-in-Chief of the "*Journal of the OR Society of China*."