Generalized Low-Rank Optimization for Topological Cooperation in Ultra-Dense Networks

Kai Yang, *Student Member, IEEE*, Yuanming Shi, *Member, IEEE*, and Zhi Ding, *Fellow, IEEE*

***Abstract*— Network densification is a natural way to support dense mobile applications under stringent requirements, such as ultra-low latency, ultra-high data rate, and massive connecting devices. Severe interference in ultra-dense networks poses a key bottleneck. Sharing channel state information (CSI) and messages across transmitters can potentially alleviate the interferences and improve the system performance. Most existing works on interference coordination require significant CSI signaling overhead and are impractical in the ultra-dense networks. This paper investigates the topological cooperation to manage interferences in message sharing based only on the network connectivity information. In particular, we propose a generalized low-rank optimization approach in a complex field to maximize the achievable degrees of freedom (DoFs) by establishing interference alignment conditions for the topological cooperation. To tackle the challenges of poor structure and non-convex rank function, we develop the Riemannian optimization algorithms to solve a sequence of *complex* fixed-rank subproblems through a rank growth strategy. By exploiting the non-compact Stiefel manifold formed by the set of complex full column rank matrices, we develop the Riemannian optimization algorithms to solve the complex fixed-rank optimization problem by applying the semi-definite lifting technique and the Burer–Monteiro factorization approach. The numerical results demonstrate the computational efficiency and higher DoFs achieved by the proposed algorithms.**

***Index Terms*— Low-rank models, topological interference alignment, transmitter cooperation, degrees-of-freedom, Riemannian optimization in complex field.**

I. Introduction

THE upsurge of wireless applications, including Internet-of-Things (IoT), Tactile Internet, tele-medicine and mobile edge artificial intelligence, is driving the paradigm shift of wireless networks from content delivery to skillset-delivery networks [1]. Network densification [2] has emerged as a promising approach to support innovative mobile applications with stringent requirements such as ultra-low latency, ultra-high data rate and massive devices connectivity. Unfortunately, interference in dense wireless network deployment becomes a key capacity limiting factor given large numbers of transmitters and receivers. Network cooperation through sharing channel state information (CSI) and messages among transmitting nodes is a viable technology to improve the spectral efficiency and energy efficiency in ultra-dense wireless networks.

Under shared CSI among transmitters, interference alignment [3] is shown to mitigate interferences base on linear coding schemes, capable of achieving half the cake for each user in K-user interference channel. Cooperative transmission [4] with message sharing has shown to be able to further improve system throughput. In particular, through centralized signal processing and interference management with full message sharing via the cloud data center, cloud radio access network (Cloud-RAN) [5] can harness the advantages of network densification. By pushing the storage resources to the network edge [6], cache-aided wireless network [7] provides a cost effective way to enable transmission cooperation.

Unfortunately, most existing works on network cooperation lead to significant channel signaling overhead. This is practically challenging in ultra-dense networks. A growing body of recent works has hence been focusing on CSI acquisition overhead reduction for interference coordination in wireless networks. Among them, delay effect in CSI acquisition has been considered in [8]. Both [9] and [10] have studied transceiver design using partial CSI, requiring instantaneous CSI for strong links and only distribution CSI of the remaining weak links. In addition, finite precision CSI feedback [11] and the compressed channel estimation [12] can further reduce CSI acquisition overhead.

However, the applicability of the aforementioned results in practical systems remains unclear, which motivates a recent proposal on topological interference management (TIM) [13]. The main idea of TIM is to manage the interference based only on the network connectivity information, which can significantly reduce the CSI acquisition overhead. By requiring only network topologies, TIM becomes one of the most promising and powerful schemes for interference management in ultra-dense wireless networks. By further enabling message sharing,

Manuscript received March 3, 2018; revised August 3, 2018 and December 7, 2018; accepted March 3, 2019. Date of publication March 26, 2019; date of current version May 8, 2019. This work was supported in part by the National Nature Science Foundation of China under Grant 61601290, in part by the Shanghai Sailing Program under Grant 16YF1407700, and in part by the National Science Foundation under Grant CNS-1702752 and Grant ECCS1711823. The associate editor coordinating the review of this paper and approving it for publication was W. Zhang. *(Corresponding author: Yuanming Shi.)*

K. Yang is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China, also with the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: yangkai@shanghaitech.edu.cn).

Y. Shi is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: shiym@shanghaitech.edu.cn).

Z. Ding is with the Department of Electrical and Computer Engineering, University of California at Davis, Davis, CA 95616 USA (e-mail: zding@ucdavis.edu).

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/TWC.2019.2904570

1536-1276 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

the work of [14] shows that transmitter cooperation based only on network topology information can strictly improve the degrees-of-freedom (DoFs). However, their results are only applicable to some specific network connectivity patterns.

In this paper, we propose a generalized low-rank optimization approach for investigating the benefits of topological cooperation for any network topology. We begin by first establishing the generalized interference alignment conditions based only on the network connectivity information with message sharing among transmitters. A low-rank model is further developed to maximize the achievable DoFs by exploiting the relationship between the model matrix rank and the achievable DoFs. The developed low-rank matrix optimization model thus generalizes the low-rank matrix completion model [15] without message sharing among transmitters. Unfortunately, the resulting generalized low-rank optimization problem in complex field is non-convex and highly intractable due to poor structure, for which novel and efficient algorithms need to be developed.

Low-rank matrix optimization models have wide range of applications in machine learning, high-dimensional statistics, signal processing and wireless networks [15]–[18]. A wealth of recent works focus on both convex approximation and non-convex algorithms to solve the non-convex and highly intractable low-rank optimization problems. Nuclear norm is a well-known convex proxy for non-convex rank function with optimality guarantees under statistical models [19]. To further reduce the storage and computation overhead for low-rank optimization, non-convex approach based on matrix factorization shows good promises [20]. With suitable statistical models, the non-convex methods can also find globally optimal solution for some structured optimization problems such as matrix completion [21]. In particular, the work of [22] adopted an alternating minimization algorithm to exploit topological transmitter cooperation gains. This algorithm stores the iterative results in the factored form and optimizes over one factor while fixing the other.

Nevertheless, the nuclear norm based convex relaxation approach in fact fails to solve the formulation of generalized low-rank matrix optimization problem because of the poor structures. Actually, the nuclear norm minimization approach always yields a full-rank matrix solution. Alternating minimization [20] algorithm by factorizing the fixed-rank matrix is particularly useful when the resulting problem is biconvex with respect to the two factors in matrix factorization. However, the convergence of the alternating minimization algorithm heavily depends on the initial points with slow convergence rates. It may also yield poor performance in achievable DoFs, as it only guarantees convergence to the first-order stationary points [15], [22]. In contrast, Riemannian optimization [23] approach has shown to be effective in improving the achievable DoFs by solving the low-rank matrix optimization problems, as the Riemannian trust-region algorithm guarantees convergence to the second-order stationary points with high precision solutions [15]. Furthermore, the Riemannian optimization algorithms are robust to initial points in ensuring convergence [24] with fast convergence rates. However, no available Riemannian optimization algorithms have been developed for the general non-square low-rank problems in the *complex field*. In this work, we develop Riemannian optimization algorithms for solving the presented generalized low-rank optimization problem in the complex field.

A. Contributions

In this paper, we develop a generalized interference alignment condition to enable transmitter cooperation based only on the network topology information. We present a generalized low-rank model to maximize the achievable DoFs. To address the special challenges in the resulting generalized low-rank optimization problem, we develop Riemannian optimization algorithms by exploiting the non-compact Stiefel manifold of fixed-rank matrices in complex field. Specifically, we propose to solve the generalized low-rank optimization problem by solving a sequence of fixed rank subproblems with rank increase. By applying semidefinite lifting technique [25], the fixed rank subproblem is reformulated as a positive semidefinite matrix problem in complex field with rank constraint. By applying the Burer-Monteiro [26] parameterization approach to factorize the positive semidefinite matrix, the resulting problem turns out to be a Riemannian optimization problem on complex non-compact Stiefel manifold. Therefore, the generalized low-rank optimization problem can be successfully solved by developing Riemannian optimization algorithms on the complex-valued non-compact Stiefel manifold.

We summarize the main contributions of this work as follows:

1) We establish a generalized interference alignment condition to enable transmitter cooperation with message sharing based only on network connectivity information. We develop a generalized low-rank model to maximize the achievable DoFs.
2) We develop first-order and second-order Riemannian optimization algorithms for solving the generalized low-rank optimization problem in *complex field*. We exploit the complex compact Stiefel manifold of complex fixed-rank matrices using the semidefinite lifting and Burer-Monteiro factorization techniques.
3) Numerical results demonstrate that the proposed second-order Riemannian trust-region algorithm is able to achieve the highest DoFs with high precision second-order stationary point solutions. Furthermore, its computing time is comparable to the first-order Riemannian conjugate gradient algorithm in medium network sizes. Overall, the Riemannian algorithms show much better performance than the alternating minimization algorithm.

B. Organization and Notation

The remainder of this paper is organized as follows. In Section II, we first introduce the system model, before establishing the generalized topological interference alignment conditions. We develop the generalized low-rank model in Section III, and derive a positive semidefinite reformulation

with the Burer-Monteiro approach to address the low-rank optimization problem in complex field. We derive Riemannian algorithms on complex non-compact Stiefel manifold in Section IV. Section V provides simulation results. Finally, we conclude this work in Section VI,

We use $[K]$ denote the set $\{1, 2, \cdots, K\}$. $\mathbb{S}_+^N$ denotes the set of all $N \times N$ Hermitian positive semidefinite matrices. And $\langle \cdot, \cdot \rangle$ denotes inner product, i.e., $\langle \boldsymbol{A}, \boldsymbol{X} \rangle = \mathrm{Tr}(\boldsymbol{A}^{\mathsf{H}} \boldsymbol{X})$.

II. System Model and Problem Formulation

In this section, we establish the generalized interference alignment condition for partially connected K-user interference channel with transmitter cooperation.

A. System Model

Consider a partially-connected interference channel with K single-antenna transmitters and K single-antenna receivers. Transmitters aim to deliver a set of independent messages $W_1, W_2, \cdots, W_K$ to receivers $1, 2, \cdots, K$, respectively. Transmitter k has message W_k and is always connected with receiver k. The channel coefficient $h_{kl} \in \mathbb{C}$ between the l-th transmitter and the k-th user is nonzero only for $(k, l) \in \mathcal{E}$. Block fading channel model is considered in this paper, i.e., h_{kl} remains stationary in r consecutive channel uses, during which the input-output relationship is given by

$$\boldsymbol{y}_k = \sum_{(k,i)\in\mathcal{E}} h_{ki} \boldsymbol{x}_i + \boldsymbol{z}_k, \quad \forall k \in [K], \tag{1}$$

where $\boldsymbol{x}_i \in \mathbb{C}^r$ is the transmitted signal at transmitter i, $\boldsymbol{y}_k \in \mathbb{C}^r$ is the received signal at receiver k, and $\boldsymbol{z}_k \in \mathbb{C}^r$ is the additive isotropic white Gaussian noise, i.e., $\boldsymbol{z}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_k)$ with $\boldsymbol{\Sigma}_k \in \mathbb{C}^{r\times r}$. Partial connectivity of the interference channel provides opportunities to enable cooperative transmission based only on the network connectivity information. Specifically, transmitter cooperation is enabled with message sharing, for which we denote the index set of messages available at transmitter k as $\mathcal{S}_k \subseteq [K]$. A 5-user example of such system is shown in Fig. 1.

Let $R(W_k)$ be the achievable data rate of message W_k, i.e., there exists a coding scheme such that the rate of message W_k is $R(W_k)$ and the decoding error probability can be arbitrarily small. Let SNR denote the signal-to-noise-ratio. For each message delivery, the degree-of-freedom (DoF) [13], the first order characterization of channel capacity, is defined as

$$\mathrm{DoF}(W_k) = \lim_{\mathrm{SNR}\to\infty} \frac{R(W_k)}{\log(\mathrm{SNR})}, \quad \forall k \in [K]. \tag{2}$$

The set of achievable DoF allocation is denoted as $\{\mathrm{DoF}(W_1), \cdots, \mathrm{DoF}(W_K)\}$, whose closure is called the DoF region. This paper adopts DoF as the performance metric and designs a linear coding scheme.

B. Linear Coding Strategy

Linear coding scheme is attractive for interference management owing to its low complexity. Specifically, its optimality in terms of DoF has been shown via interference alignment [3]. Its effectiveness has been demonstrated in the problems of topological interference management (TIM) and index coding [13]. We thus focus on linear coding scheme to design low complexity and efficient approaches for maximizing achievable DoFs.

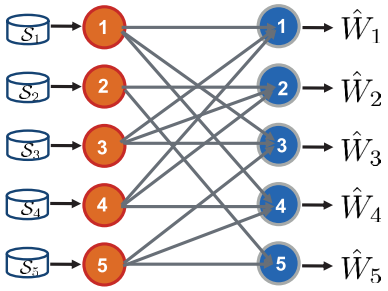


Fig. 1. The architecture of the partially-connected K-user interference channel with transmitter cooperation. $\mathcal{S}_i$ denotes the index set of messages available at transmitter i.

Suppose each message W_k is represented by a complex vector $\boldsymbol{s}_k \in \mathbb{C}^{d_k}$ with d_k data streams. Let $\boldsymbol{V}_{kj} \in \mathbb{C}^{r\times d_j}$ be the precoding matrix at transmitter k for message W_j. Then the transmitted signal is given by

$$\boldsymbol{x}_k = \sum_{j\in\mathcal{S}_k} \boldsymbol{V}_{kj} \boldsymbol{s}_j. \tag{3}$$

Consequently, the received signal at user k is

$$\boldsymbol{y}_k = \sum_{j:(k,j)\in\mathcal{E}, k\in\mathcal{S}_j} h_{kj} \boldsymbol{V}_{jk} \boldsymbol{s}_k + \sum_{i\neq k} \sum_{j:(k,j)\in\mathcal{E}, i\in\mathcal{S}_j} h_{kj} \boldsymbol{V}_{ji} \boldsymbol{s}_i + \boldsymbol{z}_k. \tag{4}$$

We let $\boldsymbol{U}_k \in \mathbb{C}^{r\times d_k}$ be the decoding matrix at receiver k.

In densified wireless networks, interference is a key bottleneck to support high data rate and low latency. To alleviate interferences by aligning the intersection of interference spaces, the following interference alignment conditions were presented in [13] and [27]

$$\sum_{j:(k,j)\in\mathcal{E}, k\in\mathcal{S}_j} h_{kj} \boldsymbol{U}_k^{\mathsf{H}} \boldsymbol{V}_{jk} \neq 0, \quad \forall k \in [K], \tag{5}$$

$$\sum_{j:(k,j)\in\mathcal{E}, i\in\mathcal{S}_j} h_{kj} \boldsymbol{U}_k^{\mathsf{H}} \boldsymbol{V}_{ji} = 0, \quad i \neq k. \tag{6}$$

Correspondingly, the message at receiver k is decoded via

$$\hat{\boldsymbol{s}}_k = \Big(\sum_{j:(k,j)\in\mathcal{E}, k\in\mathcal{S}_j} h_{kj} \boldsymbol{U}_k^{\mathsf{H}} \boldsymbol{V}_{jk} \Big)^{-1} \boldsymbol{U}_k^{\mathsf{H}} \boldsymbol{y}_k. \tag{7}$$

If there exists $\boldsymbol{U}_k$'s, $\boldsymbol{V}_{ji}$'s satisfying interference alignment conditions (5) and (6), DoF tuple $(d_1 r^{-1}, \cdots, d_K r^{-1})$ is then achievable. We thus can achieve the highest DoF by finding the minimal channel use number r.

C. Topology-Based Alignment Condition

Note that equations (5) and (6) are always feasible by increasing r. However, the interference alignment conditions (5) and (6) require the knowledge of channel coefficients h_{ij}s at the transmitters. In practice, obtaining dense network channel state information (CSI) at transmitters often requires large signaling overhead, which presents a severe obstacle to

their application in densified wireless networks. One desirable way to address the CSI acquisition overhead issue is to establish new interference alignment conditions based only on the network connectivity information, for which we present the following generalized interference alignment conditions for *topological cooperation*

$$\det\left(\sum_{j:(k,j)\in\mathcal{E},k\in\mathcal{S}_j}\boldsymbol{U}_k^{\mathsf{H}}\boldsymbol{V}_{jk}\right)\neq 0,\quad \forall k\in[K], \tag{8}$$

$$\boldsymbol{U}_k^{\mathsf{H}}\boldsymbol{V}_{ji}=\boldsymbol{0},\quad i\in\mathcal{S}_j,\ i\neq k,(k,j)\in\mathcal{E}. \tag{9}$$

Here, "topological cooperation" refers to the fact that for cooperation enabled transmitters, we design transceivers to manage interferences based on network topology information instead of instantaneous channel state information.

Proposition 1: For generic channel coefficients h_{ij}'s randomly distributed according to some continuous probability distribution [28], if (8) and (9) hold for some $\boldsymbol{U}_k, \boldsymbol{V}_{ji}$s based only on the network topology information, then they shall satisfy the channel dependent interference alignment conditions (5) and (6) with probability 1.

Proof: Let $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_T\in\mathbb{C}^{d\times d}$ denote the set of matrices $\{\boldsymbol{U}_k^{\mathsf{H}}\boldsymbol{V}_{jk}:(k,j)\in\mathcal{E},k\in\mathcal{S}_j\}$ given k. Our goal is to prove that if $\det\left(\sum_{t=1}^T\boldsymbol{Z}_t\right)\neq 0$, the probability of $\det\left(\sum_{t=1}^T h_t\boldsymbol{Z}_t\right)\neq 0$ is 1 for generic $h_1,\cdots,h_T$. Note that $\sum_{t=1}^T\boldsymbol{Z}_t=\begin{bmatrix}\boldsymbol{Z}_1&\cdots&\boldsymbol{Z}_T\end{bmatrix}\begin{bmatrix}\boldsymbol{I}&\cdots&\boldsymbol{I}\end{bmatrix}^H$. Thus, the condition $\det\left(\sum_{t=1}^T\boldsymbol{Z}_t\right)\neq 0$ implies that $\begin{bmatrix}\boldsymbol{Z}_1&\cdots&\boldsymbol{Z}_T\end{bmatrix}$ has full rank, i.e., the dimension of span$\{\boldsymbol{Z}_j\}$ is r. Since the solution to the determinant equation $\det\left(\sum_{t=1}^T h_t\boldsymbol{Z}_t\right)=0$ is an algebraic hypersurface [27], the probability of the linear combination $\sum_{t=1}^T h_t\boldsymbol{Z}_t$ with generic coefficients $h_1,\cdots,h_T$ lying on the algebraic hypersurface is hence zero. Therefore, $\det\left(\sum_{t=1}^T h_t\boldsymbol{Z}_t\right)\neq 0$ holds with probability 1. □

By leveraging conditions (8) and (9), interferences can be aligned based only on network topology CSI instead of full CSI. This significantly reduces the overhead of CSI acquisition. In particular, the topological interference alignment condition without message sharing is given by [15]

$$\det\left(\boldsymbol{U}_k^{\mathsf{H}}\boldsymbol{V}_k\right)\neq 0,\quad \forall k\in[K], \tag{10}$$

$$\boldsymbol{U}_k^{\mathsf{H}}\boldsymbol{V}_j=\boldsymbol{0},\quad j\neq k,(k,j)\in\mathcal{E}, \tag{11}$$

which is a special case of (8) and (9) with $\mathcal{S}_j=\{j\}$ and $\boldsymbol{V}_j$ denoting as $\boldsymbol{V}_{jj}$. Conditions (8) and (9) thus manifest the benefits of transmitter cooperation, as solutions to (10) and (11) are always solutions to (8) and (9), but not conversely.

Remark 1: This work assumes that there are equal number of transmitters and receivers. Nevertheless, the principle applies for arbitrary number of transmitters and receivers. This is because both Proposition 1 and the low-rank matrix representation for precoding and decoding matrices in Section 3 hold for any number of transmitters and receivers. For simplicity of notation we consider a system with K transmitters and receivers in this paper.

III. Generalized Low-Rank Optimization for Topological Cooperation

This section develops a generalized low-rank optimization framework to maximize achievable DoFs under topological cooperation. To address the challenges of the present generalized low-rank optimization problem in complex field and to exploit the algorithmic benefits of Riemannian optimization, we propose to reformulate an optimization problem over the complex non-compact Stiefel manifold by using the semidefinite lifting and Burer-Monteiro approaches.

A. Generalized Low-Rank Model for Topological Cooperation

Without loss of generality, we restrict $\sum_{j:(k,j)\in\mathcal{E},k\in\mathcal{S}_j}\boldsymbol{U}_k^{\mathsf{H}}\boldsymbol{V}_{jk}=\boldsymbol{I}$ in condition (8). By letting $m=\sum_k d_k,\ n=K\sum_k d_k$ and defining

$$\begin{aligned}\boldsymbol{U}&=\begin{bmatrix}\boldsymbol{U}_1&\cdots&\boldsymbol{U}_K\end{bmatrix}\in\mathbb{C}^{r\times m}\\ \boldsymbol{V}_j&=[\boldsymbol{V}_{j1},\cdots,\boldsymbol{V}_{jK}]\in\mathbb{C}^{r\times m}\\ \boldsymbol{V}&=\begin{bmatrix}\boldsymbol{V}_1&\cdots&\boldsymbol{V}_K\end{bmatrix}\in\mathbb{C}^{r\times n}\\ \boldsymbol{X}&=[\boldsymbol{X}_{kj}^i]=[\boldsymbol{U}_k^{\mathsf{H}}\boldsymbol{V}_{ji}]=\boldsymbol{U}^{\mathsf{H}}\boldsymbol{V}\in\mathbb{C}^{m\times n},\end{aligned} \tag{12}$$

the rank of matrix $\boldsymbol{X}$ is given as

$$\text{rank}(\boldsymbol{X})=r=d_k/\text{DoF}(W_k). \tag{13}$$

We thus can maximize the achievable DoF for interference-free message delivery by solving the following *generalized low-rank optimization* problem

$$\begin{aligned}\mathscr{P}:\ &\underset{\boldsymbol{X}\in\mathbb{C}^{m\times n}}{\text{minimize}}\ \text{rank}(\boldsymbol{X})\\ &\text{subject to }\mathcal{A}(\boldsymbol{X})=\boldsymbol{b},\end{aligned} \tag{14}$$

where the affine constraint $\mathcal{A}(\boldsymbol{X})=\boldsymbol{b}$ captures

$$\sum_{j:(k,j)\in\mathcal{E},k\in\mathcal{S}_j}\boldsymbol{X}_{kj}^k=\boldsymbol{I},\quad\forall k\in[K] \tag{15}$$

$$\boldsymbol{X}_{kj}^i=\boldsymbol{0},\quad i\neq k,\ i\in\mathcal{S}_j,\ (k,j)\in\mathcal{E} \tag{16}$$

and $\mathcal{A}:\mathbb{C}^{m\times n}\mapsto\mathbb{C}^l$.

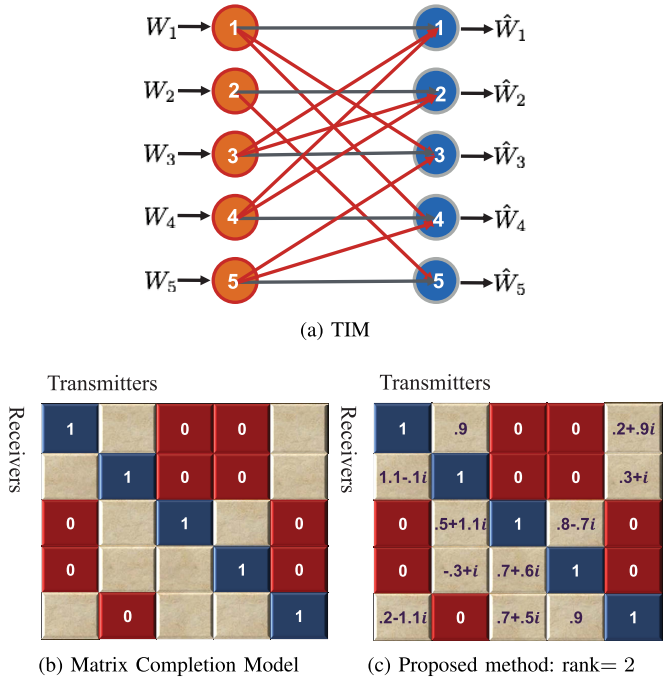
For the simpler case without message sharing, the topological interference alignment problem can be formulated as the following low-rank matrix completion problem [15], [29]

$$\begin{aligned}&\underset{\boldsymbol{X}\in\mathbb{C}^{m\times m}}{\text{minimize}}\ \text{rank}(\boldsymbol{X})\\ &\text{subject to }\boldsymbol{X}_{kk}=\boldsymbol{I},\quad\forall k\in[K]\\ &\qquad\boldsymbol{X}_{kj}=\boldsymbol{0},\quad j\neq k,\ (k,j)\in\mathcal{E},\end{aligned} \tag{17}$$

which is a special case of problem $\mathscr{P}$. The resulting low-rank matrix completion model is demonstrated in Fig. 2.

B. Problem Analysis

Basically, methods for solving low-rank problems can be divided into two categories. One uses convex relaxation approach and the other one uses nonconvex approach based on matrix factorization. In addition, penalty decomposition method is proposed in [30] for low-rank optimization problems. The inner iterations adopt a block coordinated descent



(a) TIM

(b) Matrix Completion Model

(c) Proposed method: rank= 2

Fig. 2. Matrix completion model for the topological interference alignment without message sharing for single data stream $d_k = 1$: In this case, all diagonal entries are set to 1 as $\mathcal{S}_k = \{k\}$. As an example, the (2, 3)-th entry is zero because the second receiver is connected to the interference from the third transmitter. And the (3, 2)-th entry of the matrix can be arbitrary since $(3, 2) \notin \mathcal{E}$. The solution to the low-rank optimization problem with the proposed Riemannian trust-region method is a rank-2 matrix as shown in (c) with complex entries. Therefore, 1/2 DoF is achievable for each message delivery.

method, whereas the outer iterations update the weight of rank function. However, each inner iteration requires the computation of singular value decomposition, which leads to large computation overhead ($\mathcal{O}(mnl+m^2n+m^3)$). Therefore, it is not suitable for our transceiver design problem in ultra-dense networks.

1) Convex Relaxation Methods: Nuclear norm is a well-known convex proxy [16] for rank function. The nuclear norm relaxation approach for problem $\mathscr{P}$ is given by

$$\begin{aligned} &\underset{\boldsymbol{X}\in\mathbb{C}^{m\times n}}{\text{minimize}} \ \|\boldsymbol{X}\|_* \\ &\text{subject to } \mathcal{A}(\boldsymbol{X}) = \boldsymbol{b}. \end{aligned} \tag{18}$$

It can be solved by an equivalent semidefinite programming (SDP) problem

$$\begin{aligned} &\underset{\boldsymbol{X},\boldsymbol{W}_1,\boldsymbol{W}_2}{\text{minimize}} \ \text{Tr}(\boldsymbol{W}_1) + \text{Tr}(\boldsymbol{W}_2) \\ &\text{subject to } \mathcal{A}(\boldsymbol{X}) = \boldsymbol{b}, \\ &\qquad \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{X} \\ \boldsymbol{X}^{\mathsf{H}} & \boldsymbol{W}_2 \end{bmatrix} \succeq \boldsymbol{0}. \end{aligned} \tag{19}$$

Unfortunately, the SDP solution requires computing singular value decomposition at each iteration, which is not scalable to large problem sizes in ultra-dense networks. Specifically, with high precision second-order interior point method, the convergence rate is fast while the computational cost for each iteration is $\mathcal{O}((mn + l)^3)$ due to computing the Newton step [31]. Using the first-order algorithm alternating direction method of multipliers (ADMM) [32], the computational cost is $\mathcal{O}(mnl+m^2n+m^3)$ at each iteration. Furthermore, the nuclear norm relaxation approach always yields a full rank solution due to the poor structure of the affine operator $\mathcal{A}$.

Proposition 2: The nuclear norm relaxation approach (19) for the generalized low-rank optimization problem $\mathscr{P}$ always yields a full rank solution.

Proof: See Appendix A. □

Therefore, the nuclear norm relaxation based approach is inapplicable for the poorly structured low-rank optimization problem $\mathscr{P}$. We thus call problem $\mathscr{P}$ as the generalized low-rank optimization problem.

2) Nonconvex Approaches: A rank r matrix $\boldsymbol{X}$ can be factorized as $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{R}^{\mathsf{H}}$, where $\boldsymbol{L} \in \mathbb{C}^{m\times r}$ and $\boldsymbol{R} \in \mathbb{C}^{n\times r}$. Nonconvex approaches to low-rank optimization leverage matrix factorizations and design various updating strategies for two factors $\boldsymbol{U}$ and $\boldsymbol{V}$. By solving a sequence of the fixed rank least square subproblems based on matrix factorization

$$\begin{aligned} &\underset{\boldsymbol{X}\in\mathbb{C}^{m\times n}}{\text{minimize}} \ f_0(\boldsymbol{X}) = \frac{1}{2}\|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{b}\|_2^2 \\ &\text{subject to } \text{rank}(\boldsymbol{X}) = r, \end{aligned} \tag{20}$$

and increasing r, we can find the minimal rank r for the original problem $\mathscr{P}$.

Specifically, for the rank constrained problem (20) with convex objective function, the alternating minimization [20] algorithm can function as follows:

$$\boldsymbol{L}_{k+1} = \arg\min_{\boldsymbol{L}} \ f_0(\boldsymbol{L}\boldsymbol{R}_k^{\mathsf{H}}), \tag{21}$$

$$\boldsymbol{R}_{k+1} = \arg\min_{\boldsymbol{R}} \ f_0(\boldsymbol{L}_{k+1}\boldsymbol{R}^{\mathsf{H}}). \tag{22}$$

It essentially optimizes the bi-convex objective function $f_0(\boldsymbol{L}\boldsymbol{R}^{\mathsf{H}})$ by freezing one of $\boldsymbol{L}$ and $\boldsymbol{R}$ alternatively. However, the convergence of the alternating minimization algorithm are sensitive to initial points and its convergence rate can be slow. Furthermore, it may yield poor performance for achievable DoFs maximization by only converging to first-order stationary point.

In contrast, Riemannian optimization algorithms are capable of updating the two factors $\boldsymbol{L}$ and $\boldsymbol{R}$ *simultaneously* by exploiting the quotient manifold geometry of fixed-rank matrices based on matrix factorization. First-order Riemannian conjugate gradient and second-order Riemannian trust-region algorithm can help find first-order stationary points and second-order stationary points, respectively. It has been shown in [24] that Riemannian optimization algorithms converge to first-order and second-order stationary points from arbitrary initial points. Furthermore, Riemannian trust-region algorithm can achieve high achievable DoFs with second-order stationary points while also enjoys locally super-linear [23] convergence rates.

Remark 2: The invariance of matrix factorization $(\boldsymbol{L}\boldsymbol{M}^{\mathsf{H}}, \boldsymbol{M}^{-1}\boldsymbol{R})$ for any full rank matrix $\boldsymbol{M}$ means that the critical points of the objective function f_0 parameterized with $\boldsymbol{L}$ and $\boldsymbol{R}$ are not isolated in Euclidean space. This indeterminacy also leads to an ill-conditioned Hessian [33] of f_0, which profoundly affects the convergence of second-order optimization algorithms [34], [35]. To address this issue,

we shall develop efficient algorithms on the quotient manifold instead of Euclidean space.

Unfortunately, available Riemannian algorithms for non-square fixed-rank matrix optimization problems only operate in $\mathbb{R}$ field [23] and do not directly apply to solve problem (20) in the complex filed. Inspired by the fact that the complex non-compact Stiefel manifold is well defined [36], we propose to reformulate complex matrix optimization problem (20) on the complex non-compact Stiefel manifold by using the Burer-Monteiro approach. Specifically, applying semidefinite lifting, the original problem (20) is equivalently reformulated into rank constrained positive semidefinite matrix optimization, before factorizing the semidefinite matrices using the Burer-Monteiro approach. The original complex matrix optimization problem (20) is thus reformulated as the Riemannian optimization problem over the well-defined non-compact Stiefel manifold in complex field.

C. Semidefinite Lifting and the Burer-Monteiro Approach

Burer-Monteiro approach is a well-known nonconvex parameterization method for solving positive semidefinite (PSD) matrices problems [26]. A rank r PSD matrix $\boldsymbol{Z} \in \mathbb{S}^N$ can be factorized as $\boldsymbol{Z} = \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$ with $\boldsymbol{Y} \in \mathbb{C}^{N\times r}$. The linear operator $\mathcal{A}$ can be represented as a set of matrices $\boldsymbol{A}_i \in \mathbb{C}^{n\times m}$, i.e.,

$$\mathcal{A}(\boldsymbol{X}) = [\langle \boldsymbol{A}_i, \boldsymbol{X}\rangle], \quad i = 1, \cdots, l. \tag{23}$$

Then the objective function in (20) can be rewritten as

$$f_0(\boldsymbol{X}) = \frac{1}{2}\sum_{i=1}^{l} |\langle \boldsymbol{A}_i, \boldsymbol{X}\rangle - b_i|^2. \tag{24}$$

By semidefinite lifting [25] $\boldsymbol{X}$ to

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_{11} & \boldsymbol{Z}_{12} \\ \boldsymbol{Z}_{21} & \boldsymbol{Z}_{22} \end{bmatrix} := \begin{bmatrix} \boldsymbol{L}\boldsymbol{L}^{\mathsf{H}} & \boldsymbol{L}\boldsymbol{R}^{\mathsf{H}} \\ \boldsymbol{R}\boldsymbol{L}^{\mathsf{H}} & \boldsymbol{R}\boldsymbol{R}^{\mathsf{H}} \end{bmatrix}, \tag{25}$$

we can reformulate problem (20) as a complex PSD matrix problem with rank constraint:

$$\begin{aligned} &\underset{\boldsymbol{Z}\in\mathbb{S}_+^N}{\text{minimize}} \ \ \frac{1}{2}\|\mathcal{B}(\boldsymbol{Z}) - \boldsymbol{b}\|_2^2 \\ &\text{subject to } \operatorname{rank}(\boldsymbol{Z}) = r, \end{aligned} \tag{26}$$

where $N = m + n$ and

$$\mathcal{B}(\boldsymbol{Z}) = \mathcal{A}(\boldsymbol{Z}_{12}) = \mathcal{A}(\boldsymbol{L}\boldsymbol{R}^{\mathsf{H}}). \tag{27}$$

Here we use $\mathcal{B}$ to denote a set of matrices $\boldsymbol{B}_i \in \mathbb{C}^{N\times N}$

$$\boldsymbol{B}_i = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{A}_i \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \quad \langle \boldsymbol{B}_i, \boldsymbol{Z}\rangle = \langle \boldsymbol{A}_i, \boldsymbol{X}\rangle. \tag{28}$$

We define $\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{R} \end{bmatrix} \in \mathbb{C}^{N\times r}$. The search space $\{\boldsymbol{Z} : \boldsymbol{Z} \in \mathbb{S}_+^N, \operatorname{rank}(\boldsymbol{Z}) = r\}$ admits a well-defined manifold structure, by factorizing $\boldsymbol{Z} = \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$ based on the principles of Burer-Monteiro approach. Problem (26) thus can be transformed as

$$\underset{\boldsymbol{Y}\in\mathbb{C}_*^{N\times r}}{\text{minimize}} \ f(\boldsymbol{Y}) = \frac{1}{2}\|\mathcal{B}(\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}) - \boldsymbol{b}\|_2^2. \tag{29}$$

This is a Riemannian optimization problem with a smooth (C^∞) objective function over the complex *non-compact Stiefel manifold* $\mathbb{C}_*^{N\times r}$, i.e., the set of all $N \times r$ full column rank matrices in complex field.

Algorithm 1 Optimization Framework for Transmitter Cooperation Based on Network Topology Information

Input: $\{S_j\}, \mathcal{E}, K, \{d_k\}$, accuracy ε.
Construct $\mathcal{B}$ and $\boldsymbol{b}$ following (15) (16) (23) (28). Let $N = m + n = (K+1)\sum_k d_k$.
for $r = 1, \cdots, N$ **do**
  Solve (29) with Riemannian optimization algorithm.
  if $f(\boldsymbol{Y}^{[r]}) < \varepsilon$ **then**
    return $\boldsymbol{Y}^{[r]}$
  end
end
Output: $\boldsymbol{Y}^{[r]}$ and rank r.

In summary, we propose to solve the generalized low-rank optimization problem by solving a sequence of *complex* fixed-rank optimization problem using the Riemannian optimization technique. This is achieved by lifting the complex fixed-rank optimization problem into the complex positive semidefinite matrix optimization problem, followed by parameterizing it using the Burer-Monteior approach. This yields the Riemannian optimization problem over complex non-compact Stiefel manifold. After obtaining a solution $\boldsymbol{Y}$ from (29), we can recover the solution $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{R}^{\mathsf{H}}$ to the original problem (20). The whole algorithm of addressing the transmitter cooperation problem based only on the network topology information is demonstrated in Algorithm 1.

Remark 3: We observe that the dense network connectivity (i.e., $|\mathcal{E}|$ is large) leads to a high rank r through numerical experiments in Section V-B, which results in high computational complexity if we apply the rank increase strategy in the proposed Algorithm 1. Therefore, when the number of connected links $|\mathcal{E}|$ is small, we can adopt the inflation procedure by increasing r from 1 to N. If the network connectivity is dense, we may adopt a deflation procedure by decreasing r from N to 1 to lower the complexity. For general network connectivity scenarios, a bi-section search procedure is preferred for acceleration.

IV. Matrix Optimization on Complex Non-Compact Stiefel Manifold

In this section, we shall develop Riemannian conjugate gradient and Riemannian trust-region algorithms for solving problem (29). Riemannian optimization generalizes the concepts of gradient and Hessian in Euclidean space to Riemannian gradient and Hessian on manifolds. They are represented in the tangent space, which is the linearization of the search space.

A. Quotient Geometry of Fixed-Rank Problem

For problem (29), the optima are not isolated because of $\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$ remains invariant under the *canonical projection* [23, Sec. 3.4.1]

$$\pi : \ \boldsymbol{Y} \mapsto \boldsymbol{Y}\boldsymbol{Q} \tag{30}$$

for any unitary matrix $\boldsymbol{Q} \in \mathcal{U}(r)$ where $\mathcal{U}(r)$ denotes the set of $r \times r$ unitary matrices. To address this non-uniqueness we

consider problem (29) over the equivalent class

$$[\boldsymbol{Y}] = \{\boldsymbol{YQ} : \boldsymbol{Y} \in \overline{\mathcal{M}} = \mathbb{C}_*^{N\times r}, \boldsymbol{Q} \in \mathcal{U}(r)\}, \tag{31}$$

that is

$$\underset{[\boldsymbol{Y}]\in\mathcal{M}}{\text{minimize}} \ f([\boldsymbol{Y}]). \tag{32}$$

Then the whole set of feasible solutions can be represented by isolated points in the *quotient manifold*, i.e. $\mathcal{M} = \overline{\mathcal{M}}/\sim := \overline{\mathcal{M}}/\mathcal{U}(r)$ with *canonical projection* [23, Sec. 3.4] π. Here $\sim$ is the equivalence relation and $\overline{\mathcal{M}}/\sim := \{[\boldsymbol{Y}] : \boldsymbol{Y} \in \overline{\mathcal{M}}\}$. $\overline{\mathcal{M}}$ is considered as an abstract manifold.

B. Riemannian Ingredients for Iterative Algorithms on Riemannian Manifolds

By studying the unconstrained problem on the quotient manifold instead of the constrained problem in Euclidean space, Riemannian optimization can exploit the non-uniqueness of matrix factorization with Burer-Monteiro approach. We now develop conjugate gradient and trust-region algorithms on the Riemannian manifold. To achieve this goal, we first linearize the search space, by defining the concept of *tangent space* [23, Sec. 3.5] and associated "inner product" on the tangent space. Next, we will derive the expressions for Riemannian gradient and Riemannian Hessian in this subsection.

Specifically, tangent space $\mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}}$ is a vector space consisting of all tangent vectors to $\overline{\mathcal{M}}$ at $\boldsymbol{Y}$.

Proposition 3: The tangent space of $\overline{\mathcal{M}} = \mathbb{C}_*^{N\times r}$ at $\boldsymbol{Y}$ is given by $\mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}} = \mathbb{C}^{N\times r}$.

Proof: $\overline{\mathcal{M}}$ is an open submanifold [23, Sec. 3.5.2] of $\mathbb{C}^{N\times r}$ and hence, $\mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}} = \mathbb{C}^{N\times r}$ for all $\boldsymbol{Y} \in \overline{\mathcal{M}}$. □

In order to eliminate the non-uniqueness along the equivalent class $[\boldsymbol{Y}]$, we will decompose the tangent space into two orthogonal parts, i.e., *vertical space* and *horizontal space*. Vertical space $\mathcal{V}_{\boldsymbol{Y}}$ is the tangent space of equivalent class $[\boldsymbol{Y}]$, while horizontal space $\mathcal{H}_{\boldsymbol{Y}}$ is the orthogonal complement of vertical space in the tangent space. That is,

$$\mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}} = \mathcal{V}_{\boldsymbol{Y}} \oplus \mathcal{H}_{\boldsymbol{Y}}, \tag{33}$$

where $\oplus$ denotes the direct sum of two subspace. In this way, we can always find the unique "lifted" representation of the tangent vectors of $\mathcal{T}_{[\boldsymbol{Y}]}\mathcal{M}$ in $\mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}}$ at any element of $[\boldsymbol{Y}]$, i.e., for any $\boldsymbol{\xi} \in \mathcal{T}_{\boldsymbol{Y}}\mathcal{M}$ we define a unique horizontal lift $\overline{\boldsymbol{\xi}} \in \mathcal{H}_{\boldsymbol{Y}}$ at $\boldsymbol{Y}$ such that

$$\overline{\boldsymbol{\xi}} := \Pi_{\boldsymbol{Y}}^h \boldsymbol{\xi}, \tag{34}$$

where *horizontal projection* $\Pi_{\boldsymbol{Y}}^h(\cdot)$ is the orthogonal projection from $\mathcal{T}_{\boldsymbol{Y}}\mathcal{M}$ onto $\mathcal{H}_{\boldsymbol{Y}}$.

Proposition 4: The vertical space at $\boldsymbol{Y}$ is given by

$$\mathcal{V}_{\boldsymbol{Y}} \triangleq \{\boldsymbol{Y\Omega} : \boldsymbol{\Omega}^{\mathsf{H}} = -\boldsymbol{\Omega}, \boldsymbol{\Omega} \in \mathbb{C}^{r\times r}\}. \tag{35}$$

Proof: See Appendix B. □

According to the definition, horizontal space should be derived from

$$\mathcal{H}_{\boldsymbol{Y}} = \{\overline{\boldsymbol{\xi}} \in \mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}} : \overline{g}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}}, \overline{\boldsymbol{\zeta}}) = 0, \ \forall \overline{\boldsymbol{\zeta}} \in \mathcal{V}_{\boldsymbol{Y}}\}, \tag{36}$$

where $\overline{g}$ is *Riemannian metric* for the abstract manifold $\overline{\mathcal{M}}$. Riemannian metric is the generalization of "inner product" in Euclidean space to a manifold. It is a bilinear, symmetric positive-definite operator

$$\overline{g} : \ \mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}} \times \mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}} \mapsto \mathbb{R}. \tag{37}$$

In this paper, we can choose

$$\overline{g}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}}, \overline{\boldsymbol{\zeta}}) := \mathrm{Tr}(\Re(\overline{\boldsymbol{\xi}}^{\mathsf{H}}\overline{\boldsymbol{\zeta}})) = \frac{1}{2}\mathrm{Tr}(\overline{\boldsymbol{\xi}}^{\mathsf{H}}\overline{\boldsymbol{\zeta}} + \overline{\boldsymbol{\zeta}}^{\mathsf{H}}\overline{\boldsymbol{\xi}}) \tag{38}$$

as a Riemannian metric for the abstract manifold $\overline{\mathcal{M}}$, where $\boldsymbol{Y} \in \overline{\mathcal{M}}$ and $\overline{\boldsymbol{\xi}}, \overline{\boldsymbol{\zeta}} \in \mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}}$. The manifold $\overline{\mathcal{M}}$ is called a *Riemannian manifold* when its tangent spaces are endowed with a Riemannian metric. From another perspective, $\overline{\mathcal{M}}$ can also be viewed as a Kähler manifold whose Kähler form is a real closed (1,1)-form [37].

Therefore, we can obtain the explicit expressions for the horizontal space and horizontal projection.

Proposition 5: The horizontal space is

$$\mathcal{H}_{\boldsymbol{Y}} = \{\boldsymbol{\xi} \in \mathbb{C}^{N\times r} : \boldsymbol{\xi}^{\mathsf{H}}\boldsymbol{Y} = \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\xi}\}, \tag{39}$$

and the orthogonal projection onto the horizontal space is

$$\Pi_{\boldsymbol{Y}}^h \boldsymbol{\xi}_{\boldsymbol{Y}} = \boldsymbol{\xi}_{\boldsymbol{Y}} - \boldsymbol{Y\Omega}, \tag{40}$$

where $\boldsymbol{\Omega}^{\mathsf{H}} = -\boldsymbol{\Omega} \in \mathbb{C}^{r\times r}$ is the solution of Lyapunov equation

$$\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Y\Omega} + \boldsymbol{\Omega}\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Y} = \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\xi}_{\boldsymbol{Y}} - \boldsymbol{\xi}_{\boldsymbol{Y}}^{\mathsf{H}}\boldsymbol{Y}. \tag{41}$$

Proof: See Appendix C. □

Given the Riemannian metric for the abstract manifold, quotient manifold is naturally endowed with a Riemannian metric

$$g_{[\boldsymbol{Y}]}(\boldsymbol{\xi}_{[\boldsymbol{Y}]}, \boldsymbol{\zeta}_{[\boldsymbol{Y}]}) := \overline{g}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}}_{\boldsymbol{Y}}, \overline{\boldsymbol{\zeta}}_{\boldsymbol{Y}}) \tag{42}$$

such that the expression $\overline{g}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}}_{\boldsymbol{Y}}, \overline{\boldsymbol{\zeta}}_{\boldsymbol{Y}})$ remains for any elements in the equivalent class $[\boldsymbol{Y}]$. Hence $\mathcal{M}$ is a *Riemannian quotient manifold* of the abstract manifold $\overline{\mathcal{M}}$ with the Riemannian metric g, and the canonical projection $\pi : (\overline{\mathcal{M}}, \overline{g}) \mapsto (\mathcal{M}, g)$ is a *Riemannian submersion* [23, Sec. 3.6.2].

Riemannian optimization generalizes the gradient and Hessian into Riemannian gradient and Riemannian Hessian.

1) Riemannian Gradient: Riemannian gradient is a necessary ingredient to develop the Riemannian conjugate gradient and Riemannian trust-region algorithm. For quotient manifold, the horizontal representation of Riemannian gradient, denoted by $\mathrm{grad} f(\boldsymbol{Y})$, arises from

$$\mathrm{grad} f(\boldsymbol{Y}) = \Pi_{\boldsymbol{Y}}^h \overline{\mathrm{grad}} f(\boldsymbol{Y}), \tag{43}$$

in which $\overline{\mathrm{grad}} f(\boldsymbol{Y})$ is the Riemannian gradient in the abstract manifold $\overline{\mathcal{M}}$ at $\boldsymbol{Y}$. Note that $\overline{\mathrm{grad}} f(\boldsymbol{Y})$ is given by

$$\overline{g}_{\boldsymbol{Y}}(\overline{\mathrm{grad}} f(\boldsymbol{Y}), \overline{\boldsymbol{\xi}}) = \mathrm{D}f(\boldsymbol{Y})[\overline{\boldsymbol{\xi}}], \quad \forall \overline{\boldsymbol{\xi}} \in \mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}}, \tag{44}$$

where $\mathrm{D}f(\boldsymbol{Y})[\boldsymbol{\xi}] := \lim_{t\to 0} t^{-1}\left[f(\boldsymbol{Y} + t\boldsymbol{\xi}) - f(\boldsymbol{Y})\right]$ is the directional derivative of f, whereas $\overline{\boldsymbol{\xi}}$ is the horizontal lift of $\boldsymbol{\xi}$. Then we conclude that

$$\mathrm{grad} f(\boldsymbol{Y}) = \overline{\mathrm{grad}} f(\boldsymbol{Y}) = \sum_{i=1}^{l} (C_i \boldsymbol{B}_i + C_i^* \boldsymbol{B}_i^{\mathsf{H}})\boldsymbol{Y}, \tag{45}$$

in which $C_i = \langle \boldsymbol{B}_i, \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}} \rangle - b_i$. The derivation process is described in detail in Appendix D.

TABLE I
RIEMANNIAN INGREDIENTS

$\text{minimize}_{[\boldsymbol{Y}]\in\overline{\mathcal{M}}}\ f(\boldsymbol{Y})$		
Computation space	$\overline{\mathcal{M}} = C_*^{N\times r}, \mathcal{M} = \overline{\mathcal{M}}/\sim$	
Canonical projection	$\pi : \boldsymbol{Y} \mapsto \boldsymbol{YQ}, \boldsymbol{Q} \in \mathcal{U}(r)$	(30)
Remannian metric	$g_{\boldsymbol{Y}}(\boldsymbol{\xi_Y}, \boldsymbol{\eta_Y}) = \text{Tr}(\boldsymbol{\xi}_{\boldsymbol{Y}}^{\mathsf{H}}\boldsymbol{\eta_Y} + \boldsymbol{\eta}_{\boldsymbol{Y}}^{\mathsf{H}}\boldsymbol{\xi_Y})$	(38)
Vertical space	$\mathcal{V}_{\boldsymbol{Y}} = \{\boldsymbol{Y\Omega} : \boldsymbol{\Omega}^{\mathsf{H}} = -\boldsymbol{\Omega}, \boldsymbol{\Omega} \in \mathbb{C}^{r\times r}\}$	(35)
Horizontal space	$\mathcal{H}_{\boldsymbol{Y}} = \{\boldsymbol{\xi} \in \mathbb{C}^{N\times r} : \boldsymbol{\xi}^{\mathsf{H}}\boldsymbol{Y} = \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\xi}\}$	(39)
Projection onto horizontal space	$\Pi_{\mathcal{M}}^{h}\boldsymbol{\xi_Y} = \boldsymbol{\xi_Y} - \boldsymbol{Y\Omega}$	(40)
Remannian gradient	$\text{grad} f(\boldsymbol{Y}) = \sum_{i=1}^{l}(C_i\boldsymbol{B}_i + C_i^*\boldsymbol{B}_i^{\mathsf{H}})\boldsymbol{Y}$	(45)
Remannian Hessian	$\text{Hess} f(\boldsymbol{Y})[\boldsymbol{\eta_Y}]$	(48)
Retraction	$\mathcal{R}_{\boldsymbol{Y}}(\boldsymbol{\xi}) = \pi(\boldsymbol{Y} + \bar{\boldsymbol{\xi}})$	(50)

2) Riemannian Hessian: For the purpose of developing a second-order algorithm, we need to think of the Riemannian Hessian as an linear operator closely connected to the directional derivative of the gradient. *Riemannian connection* defines a "directional derivative" on the Riemannian manifold. To be specific, Euclidean directional derivative is a Riemannian connection on $\mathbb{C}^{N\times r}$. Since the quotient manifold $\mathcal{M}$ has a Riemannian metric that is invariant along the horizontal space, the Riemannian connection [23, Proposition 5.3.4] can be derived from

$$\nabla_{\boldsymbol{\eta}}\boldsymbol{\xi} = \Pi_{\boldsymbol{Y}}^{h}(\text{D}\overline{\boldsymbol{\xi}}[\overline{\boldsymbol{\eta}}]), \tag{46}$$

for any $\boldsymbol{\eta} \in \mathcal{V}_{\boldsymbol{Y}}$, $\boldsymbol{\xi} \in \mathfrak{X}(\mathcal{M})$ and $\mathfrak{X}(\mathcal{M})$ is the set of smooth vector fields on $\mathcal{M}$. The horizontal representation of Riemannian Hessian operator [23, Definition 5.5.1] is given as

$$\text{Hess} f(\boldsymbol{Y})[\boldsymbol{\xi_Y}] := \nabla_{\boldsymbol{\xi_Y}}\overline{\text{grad}} f. \tag{47}$$

Then the Riemannian Hessian is given by

$$\text{Hess} f(\boldsymbol{Y})[\boldsymbol{\eta_Y}] = \Pi_{\boldsymbol{Y}}^{h}\Bigg(\sum_{i=1}^{l}(C_{\boldsymbol{\eta}_i}\boldsymbol{B}_i\boldsymbol{Y} + C_i\boldsymbol{B}_i\boldsymbol{\eta}_Y + C_{\boldsymbol{\eta}_i}^*\boldsymbol{B}_i^{\mathsf{H}}\boldsymbol{Y} + C_i^*\boldsymbol{B}_i^{\mathsf{H}}\boldsymbol{\eta_Y})\Bigg), \tag{48}$$

where $C_{\boldsymbol{\eta}_i} = \langle \boldsymbol{B}_i, \boldsymbol{Y}\boldsymbol{\eta}_{\boldsymbol{Y}}^{\mathsf{H}} + \boldsymbol{\eta_Y}\boldsymbol{Y}^{\mathsf{H}}\rangle$. We relegate the derivation details of this expression to Appendix D.

C. Riemannian Optimization for Fixed-Rank Problem

Riemannian optimization generalizes the optimization algorithms in Euclidean space to a manifold. Similarly, we need to compute search directions in the tangent space and appropriate stepsizes. To ensure each iteration is always on the given manifold, *retraction* [23, Sec. 4.1] is defined as a pull-back from the tangent space onto the manifold. To be specific, the updating formula in the i-th iteration is given by

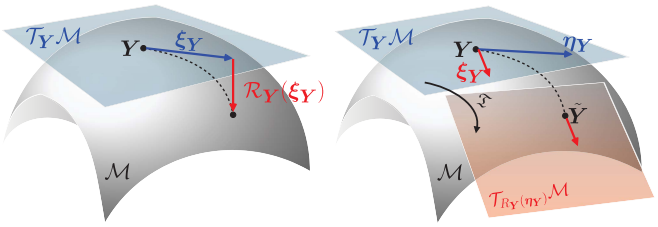
$$\boldsymbol{Y}_{k+1} = \mathcal{R}_{\boldsymbol{Y}_k}(\alpha_k\boldsymbol{\eta}_k), \tag{49}$$

where $\alpha_k > 0$ is the step size, $\boldsymbol{\eta}_k \in \mathcal{T}_{\boldsymbol{Y}_k}\mathcal{M}$ is the search direction, and $\mathcal{R}$ denotes retraction operation which maps an element from the set of all tangent spaces $\mathcal{T}\mathcal{M} = \cup_{\boldsymbol{Y}\in\mathcal{M}}\mathcal{T}_{\boldsymbol{Y}}\mathcal{M}$ to $\mathcal{M}$. The retraction operation is shown in Fig. 3a.

Proposition 6: Choices of $\overline{\mathcal{R}}$ and $\mathcal{R}$

$$\overline{\mathcal{R}}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}}) := \boldsymbol{Y} + \overline{\boldsymbol{\xi}}, \quad \mathcal{R}_{\boldsymbol{Y}}(\boldsymbol{\xi}) := \pi(\overline{\mathcal{R}}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}})) \tag{50}$$

define retractions on $\overline{\mathcal{M}}$ and $\mathcal{M}$, respectively.



(a) Retraction (b) Vector Transport

Fig. 3. Riemannian retraction and vector transport operation.

Proof: Since $\overline{\mathcal{M}} = \mathbb{C}_*^{N\times r}$ is an embedded manifold and also an open submanifold of $\mathcal{E} = \mathbb{C}^{N\times r}$, following [23, Sec. 4.1.1] we can choose the identity mapping

$$\phi(\boldsymbol{F}) = \boldsymbol{F} \tag{51}$$

as a diffeomorphism so that $\phi : \overline{\mathcal{M}} \times \overline{\mathcal{N}} \to \mathcal{E}_*, \overline{\mathcal{N}} = \emptyset$ and $\dim(\overline{\mathcal{M}}) + \dim(\overline{\mathcal{N}}) = \dim(\mathcal{E})$. Therefore, we conclude that

$$\overline{\mathcal{R}}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}}) := \boldsymbol{Y} + \overline{\boldsymbol{\xi}} \tag{52}$$

defines a retraction on $\overline{\mathcal{M}}$. Adding with that equivalent classes are orbits of the Lie group $\mathcal{U}_r$ which acts linearly [23, Sec. 4.1.2] on the abstract manifold $\overline{\mathcal{M}}$,

$$\mathcal{R}_{\boldsymbol{Y}}(\boldsymbol{\xi}) := \pi(\overline{\mathcal{R}}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}})) \tag{53}$$

defines a retraction on $\mathcal{M}$. □

In this subsection, we will introduce Riemannian conjugate gradient (RCG) method and Riemannian trust-region (RTR) method.

1) Riemannian Conjugate Gradient Method: When the search direction is chosen as the negative Riemannian gradient and the step size is determined by backtracking line search following the Armijo rule [23, Sec. 4.6.3], we have the Riemannian gradient descent algorithm. Riemannian conjugate gradient method can be expressed as

$$\boldsymbol{\eta}_{k+1} = -\text{grad} f_k + \beta_k\mathfrak{T}_{\alpha_k\boldsymbol{\eta}_k}(\boldsymbol{\eta}_k), \tag{54}$$

where $\mathfrak{T}_{\boldsymbol{\eta_Y}}(\boldsymbol{\xi_Y})$ is the *vector transport* operator so that $\boldsymbol{\xi_Y}$ is transported from $\mathcal{T}_{\boldsymbol{Y}}\mathcal{M}$ to $\mathcal{T}_{\mathcal{R}_{\boldsymbol{Y}}(\boldsymbol{\eta_Y})}\mathcal{M}$ for $\boldsymbol{\xi_Y} \in \mathcal{T}_{\boldsymbol{Y}}\mathcal{M}$. This is shown in Fig. 3b.

A vector transport is defined by

$$\mathfrak{T}_{\boldsymbol{\eta_Y}}(\boldsymbol{\xi_Y}) := D\mathcal{R}_{\boldsymbol{Y}}(\boldsymbol{\eta_Y})[\boldsymbol{\xi_Y}] = \boldsymbol{\xi_Y}. \tag{55}$$

Among many good choices for β_k, we choose

$$\beta_k = \frac{g_{\boldsymbol{Y}_k}(\text{grad} f_k, \text{grad} f_k - \mathfrak{T}_{\alpha_{k-1}\boldsymbol{\eta}_{k-1}}(\text{grad} f_{k-1}))}{g_{\boldsymbol{Y}_k}(\boldsymbol{\eta}_{k-1}, \text{grad} f_k - \mathfrak{T}_{\alpha_{k-1}\boldsymbol{\eta}_{k-1}}(\text{grad} f_{k-1}))}, \tag{56}$$

which is a generalized version of Hestenes-Stiefel [38].

Algorithm 2 Truncated Conjugate Gradient Algorithm for (57)

Input: $\mathcal{B}, \boldsymbol{b}, \boldsymbol{Y}_k, \Delta_k$. Parameters $\kappa, \theta > 0$
Initialize: $\boldsymbol{\eta}^0, \boldsymbol{r}_0 = \text{grad} f_k, \boldsymbol{\delta}_0 = -\boldsymbol{r}_0$
while $\|\boldsymbol{r}_{j+1}\|_g > \|\boldsymbol{r}_0\|_g \min(\|\boldsymbol{r}_0\|_g^{\theta}, \kappa)$ **do**
  if $g_{\boldsymbol{Y}_k}(\boldsymbol{\delta}_j, \mathit{Hess} f_k[\boldsymbol{\delta}_j]) \leq 0$ **then**
    Compute $\tau = \arg\min m_k(\boldsymbol{\eta}_k)$ where $\boldsymbol{\eta}_k = \boldsymbol{\eta}^j + \tau\boldsymbol{\delta}_j$ and $\|\boldsymbol{\eta}_k\|_g = \Delta_k$,
    return $\boldsymbol{\eta}$
  end
  Set $\boldsymbol{\eta}^{j+1} = \boldsymbol{\eta}^j + \alpha_j \boldsymbol{\delta}_j$ where $\alpha_j = \|\boldsymbol{r}_j\|_g^2 / g_{\boldsymbol{Y}_k}(\boldsymbol{\delta}_j, \text{Hess} f_k[\boldsymbol{\delta}_j])$
  if $\|\boldsymbol{\eta}^{j+1}\|_g \geq \Delta_k$ **then**
    Set τ as the solution to $\|\boldsymbol{\eta}_k\|_g = \Delta_k$ where $\boldsymbol{\eta}_k = \boldsymbol{\eta}^j + \tau\boldsymbol{\delta}_j$,
    return $\boldsymbol{\eta}_k$
  end
  Set $\boldsymbol{r}_{j+1} = \boldsymbol{r}_j + \alpha_j \text{Hess} f_k[\boldsymbol{\delta}_j]$,
  Set $\beta_{j+1} = \|\boldsymbol{r}_{j+1}\|_g^2 / \|\boldsymbol{r}_j\|_g^2$,
  Set $\boldsymbol{\delta}_{j+1} = -\boldsymbol{r}_{j+1} + \beta_{j+1}\boldsymbol{\delta}_j$,
  $j \leftarrow j + 1$
end
Output: $\boldsymbol{\eta} = \boldsymbol{\eta}_k$.

2) Riemannian Trust-Region Algorithm: When the search direction is chosen by solving the local second-order approximation of $f(\boldsymbol{Y})$, it results in the Riemannian trust-region algorithm. We will find the updating vector $\boldsymbol{\eta}$ by solving the trust-region subproblem

$$\begin{aligned} \underset{\boldsymbol{\eta} \in \mathcal{T}_{\boldsymbol{Y}_k}\mathcal{M}}{\text{minimize}} \; & m_k(\boldsymbol{\eta}) = f_k + g_{\boldsymbol{Y}_k}(\text{grad} f_k, \boldsymbol{\eta}) \\ & \qquad + \frac{1}{2} g_{\boldsymbol{Y}_k}(\text{Hess} f_k[\boldsymbol{\eta}], \boldsymbol{\eta}) \\ \text{subject to } & \|\boldsymbol{\eta}\|_g \leq \Delta_k, \end{aligned} \tag{57}$$

where $\Delta_k > 0$ is the radius of the trust region and $\|\boldsymbol{\eta}\|_g = \sqrt{g_{\boldsymbol{Y}_k}(\boldsymbol{\eta}, \boldsymbol{\eta})}$. Note that the solution $\boldsymbol{\eta}$ becomes a candidate for updating. This is because we will select a proper Δ_k, find the corresponding solution $\boldsymbol{\eta}_k$ of the trust-region subproblem and then update $\boldsymbol{Y}$ through

$$\boldsymbol{Y}_{k+1} = \mathcal{R}_{\boldsymbol{Y}_k}(\boldsymbol{\eta}_k). \tag{58}$$

The criterion for choosing Δ_k is based on evaluating

$$\rho_k = \frac{f_k - f(\mathcal{R}_{\boldsymbol{Y}_k}(\boldsymbol{\eta}_k))}{m_k(\boldsymbol{0}) - m_k(\boldsymbol{\eta}_k)}. \tag{59}$$

When ρ_k is very small, the radius of trust region Δ_k should be reduced because in this case the second-order approximation is too inaccurate. If ρ_k is not very small, we shall accept Δ_k and $\boldsymbol{\eta}_k$ and reduce the trust region. If ρ_k is close to 1, it m means that the second-order approximation models original objective function well. Hence, we can accept this step and expand the trust region. Likewise, if $\rho_k \ll 1$, we should also reject this step and increase Δ_k. The trust-region subproblem can be solved by the truncated conjugate gradient [23, Sec. 7.3.2] algorithm (see Algorithm 2).

Riemannian trust-region algorithm harnesses the second-order information of the problem. It admits a superlinear [23, Th. 7.4.11] convergence rate locally and is robust to initial points. Since the objective function f is exactly a quadratic function which satisfies the Lipschitz gradient condition and other assumptions in [24], we can always find an approximate second-order critical points by the Riemannian trust-region algorithm.

D. Computational Complexity Analysis

Riemannian conjugate gradient and Riemannian trust-region algorithm involve computing optimization ingredients at each iteration, for which we show their computational complexity.

- Evaluate the objective value $f(\boldsymbol{Y})$. Since $\mathcal{B}(\cdot)$ involves a series of sparse matrix multiplication, we can compute it efficiently and the complexity of computing $f(\boldsymbol{Y})$ is $\mathcal{O}(mnl)$.
- Compute the Riemannian gradient grad$f$ (45). This includes computing matrix multiplication in $\mathcal{O}(mnl)$ and horizontal projection $\Pi^h_{\boldsymbol{Y}}$ (40). Since complexity of solving the Lyapunov equation (41) is $\mathcal{O}(r^3 + (m+n)r^2)$, the overall complexity is $\mathcal{O}(mnl + (m+n)r^2 + r^3)$.
- Compute the Riemannian Hessian Hess$f$ (48). Its cost is also $\mathcal{O}(mnl + (m+n)r^2 + r^3)$.
- Computing the Riemannian metric $\overline{g}$ (38). This complexity is dominant by matrix multiplications, which is $\mathcal{O}((m+n)r^2)$.
- Computational complexity of retraction $\overline{\mathcal{R}}$ (50) is $\mathcal{O}((m+n)r)$ and vector transport $\mathfrak{T}$ (55) is insignificant.

From the above results, we conclude that the computational complexity of Riemannian conjugate gradient algorithm for each iteration is $\mathcal{O}(mnl + (m+n)r^2 + r^3)$. And each iteration of truncated conjugate gradient algorithm also involves computation with complexity $\mathcal{O}(mnl + (m+n)r^2 + r^3)$.

V. Simulations

This section presents numerical experiments to demonstrate the efficacy of the generalized low-rank optimization approach for topological cooperation via the newly presented Riemannian optimization algorithms. We will investigate the performance of different algorithms from the perspective of convergence rate and achievable DoF. We evaluate our model in different settings and demonstrate that the generalized low-rank approach can effectively enable transmitter cooperation based only on network topology information.

Our simulations compare the following matrix-factorization-based algorithms for solving the generalized low-rank optimization problem $\mathscr{P}$:

- "AltMin" [22]: Alternating minimization algorithm (21) (22) is adopted in [22] for topological transmitter cooperation problem by alternatively updating factors. For fixed r, (21) (22) are solved with gradient descent followed by backtracking line search.
- "RCG": The Riemannian conjugate gradient method is developed in Sec IV-C.1. We implement this algorithm with *Manopt* [39] software package.
- "RTR": The Riemannian trust-region method is developed in Sec IV-C.2 and also implemented with *Manopt*.

All algorithm are adopted with random initialization strategy for each rank r, and we find the minimal r by increasing r

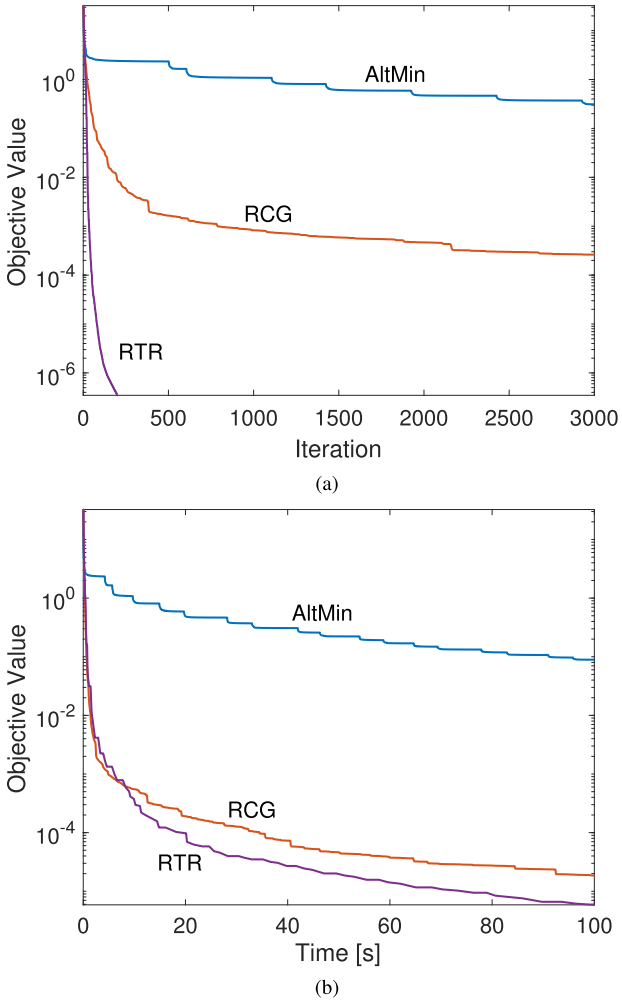


Fig. 4. Convergence rate and computing time of all algorithms with full transmitter cooperation.

from 1 to N until $m^{-0.5}\cdot\|\mathcal{A}(\boldsymbol{X})-\boldsymbol{b}\|<10^{-3}$. In our numerical experiments, the network topology and shared messages at each transmitter are generated uniformly at random with probabilities

$$\text{Prob}((k,j)\in\mathcal{E})=\begin{cases}p, & j\neq k\\ 1, & j=k,\end{cases} \tag{60}$$

and

$$\text{Prob}(j\in\mathcal{S}_k)=\begin{cases}q, & j\neq k\\ 1, & j=k,\end{cases} \tag{61}$$

respectively.

A. Convergence Rate

Consider a partially connected 20-user interference channel with full transmitter cooperation. The network topology is generated randomly and each link is connected with probability $p=0.3$. Each message is split into 3 data streams. In this simulation, $K=20, d_1=\cdots=d_K=3$, $p=0.3$ and $r=12$. Fig. 4 shows the convergence behaviors of all 3 algorithms in terms of iterations and time. The results indicate that the proposed RTR algorithm exhibits a superlinear convergence rate, and the computing rate is comparable with first order

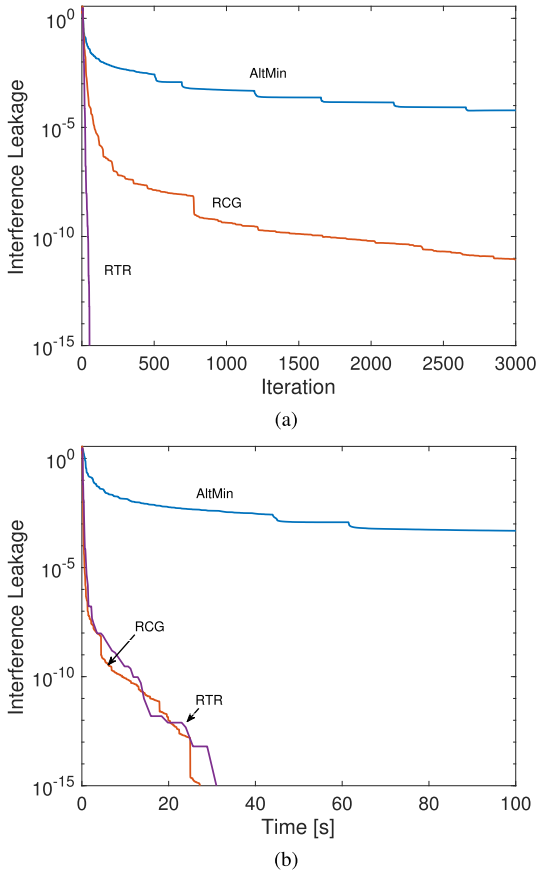


Fig. 5. Convergence of interference leakage for all algorithms with full transmitter cooperation.

RCG algorithm. In addition, the proposed RTR can yield a more accurate solution with the second-order stationary point when compared against first order algorithms that guarantee convergence only to first-order stationary points. The overall test results show that the proposed RTR and RCG algorithms are much more efficient than other contemporary algorithms in terms of convergence rates and solution performance.

To further show that the interferences are nulled, we choose the interference leakage as the metric and plot it in Fig. 5 for the same setting of Fig. 4. The interference leakage cost is given by

$$IL=\sum_{i\neq k}\sum_{j:(k,j)\in\mathcal{E},i\in\mathcal{S}_j}\|h_{kj}\boldsymbol{U}_k^{\mathsf{H}}\boldsymbol{V}_{ji}\Big(h_{kj}\boldsymbol{U}_k^{\mathsf{H}}\boldsymbol{V}_{ji}\Big)^{\mathsf{H}}\|_F^2, \tag{62}$$

where the channel coefficients follow standard complex Gaussian distribution. Fig. 5 demonstrates that there is a rapid decline of interference leakage as the objective value decreases with the proposed Riemannian optimization algorithms.

B. DoF Over Network Topologies

Consider a partially connected 20-user interference channel without message splitting ($d_k=1$). The network topologies are generated randomly with different p. Fig. 6 demonstrates the DoF over p with full transmitter cooperation. Each DoF result is averaged over 100 times. This result shows that, among the 3 solutions, the proposed RTR algorithm achieves the best performance with second-order stationary points. The

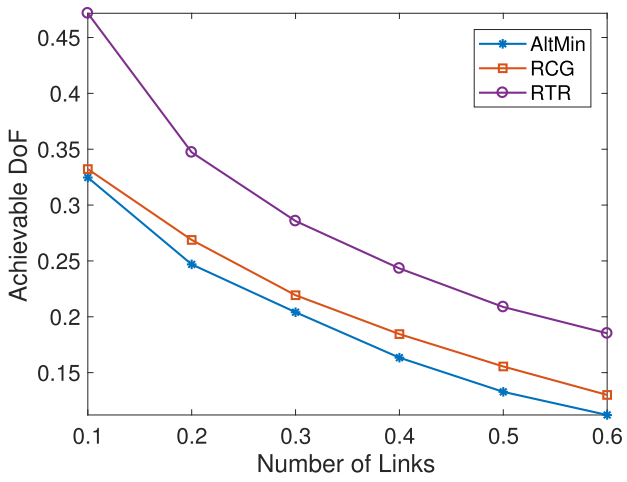


Fig. 6. DoF over the number of connected links.

Riemannian algorithms RTR and RCG significantly outperform the alternating minimization algorithm owing to their good convergence guarantee.

To further justify the effectiveness of the Riemannian optimization framework, we check the recovered DoF returned by the proposed RTR algorithm for all specific network topologies $\mathcal{E}$ and specific message sharing pattern $\{\mathcal{S}_k : k \in [K]\}$ provided in [14]. Specifically, transmitter cooperation improves the symmetric DoF from 1/3 to 2/5 for Example 1 in Fig. 1(a), from 2/5 to 1/2 for Example 4 in Fig. 3(a), and from 1/3 to 2/5 for Example 7 in Fig. 6(a), compared with the cases without cooperation. And the optimal symmetric DoF is 1/2 for Example 6 in Fig. 5(a) with transmitter cooperation. All these optimal symmetric DoF results can be achieved by the proposed RTR algorithm numerically. However, theoretically identifying the network topologies and the message sharing patterns for which the Riemannian trust region algorithm can provide optimal symmetric DoFs is still a challenging open problem.

C. Transmitter Cooperation Gains

We investigate the achievable DoFs in partially connected 20-user interference channels. We randomly generate the network topologies with $p = 0.2$ and simulate different algorithms under different transmitter cooperation level q with single data stream. For each cooperation level, we take average over 500 channel realizations. Fig. 7a shows that the second-order algorithm RTR can achieve the highest DoF among all algorithms. Comparing the first-order algorithms, the proposed RCG outperforms AltMin. With the high convergence rate and second-order stationary points solutions of RTR, the gap between RTR algorithm and other algorithms grows with q, which indicates that the proposed RTR algorithm is capable of fully leveraging the benefits of transmitter cooperation.

To further illustrate the transmitter cooperation benefit, we evaluate the achievable sum-rate using the proposed RTR algorithm in Fig. 7a. In this single data stream test setting, $\boldsymbol{U}_k, \boldsymbol{V}_{ji}$ degenerates to vectors $\boldsymbol{u}_k, \boldsymbol{v}_{ji}$. Assume that each single data stream symbol s_i has unit power, i.e., $\mathbb{E}(|s_i|^2) = 1$. Suppose the noise is i.i.d. Gaussian, i.e., $\boldsymbol{\Sigma}_k = \sigma^2 \boldsymbol{I}_r$,

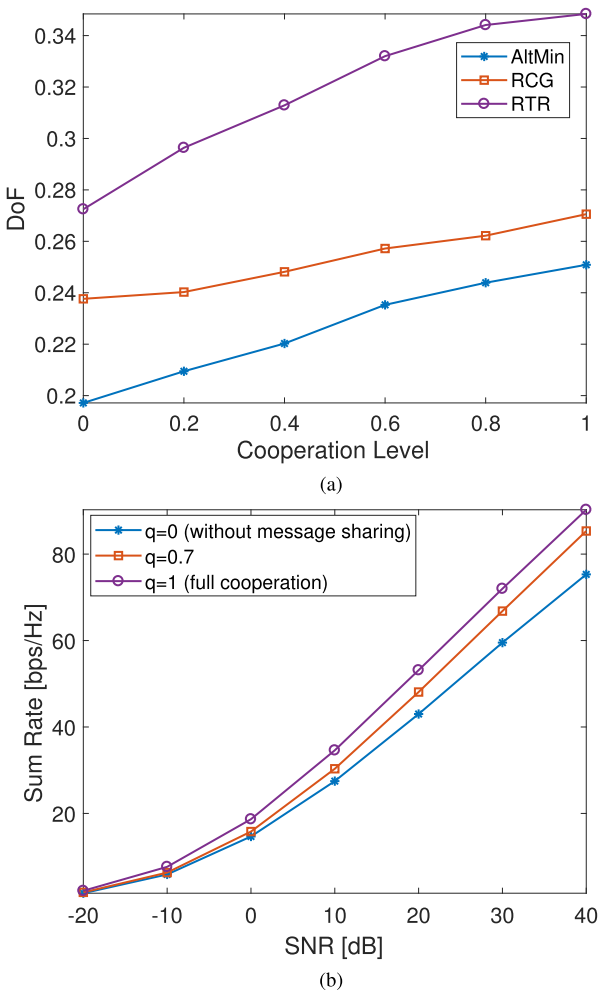


Fig. 7. Benefits of transmitter cooperation: (a) DoF over different transmitters cooperation levels q. (b) Sum-rate over the transmit power.

with σ^2 at -120dB. The distance d_{ij} between each connected transmitter-receiver pair (j, i) is uniformly distributed in $[0.1, 0.2]$ km. The fading channel model is given as

$$h_{ij} = 10^{-L(d_{ij})/20} c_{ij}, \quad (i,j) \in \mathcal{E}, \tag{63}$$

where the pass loss is given by $L(d_{ij}) = 128.1 + 37.6 \log_{10} d_{ij}$ and the small scale fading coefficient is given by $c_{ij} \sim \mathcal{CN}(0,1)$. Then the sum rate per channel use is given by

$$C_{\text{sum}} = \frac{1}{r} \sum_{k=1}^{K} C_k = \frac{1}{r} \sum_{k=1}^{K} \log(1 + \text{SINR}_k), \tag{64}$$

where

$$\text{SINR}_k = \frac{|\sum_{(k,j)\in\mathcal{E}, k\in\mathcal{S}_j} h_{kj} \boldsymbol{u}_k^{\mathsf{H}} \boldsymbol{v}_{jk}|^2}{\sum_{i\neq k} |\sum_{(k,j)\in\mathcal{E}, i\in\mathcal{S}_j} h_{kj} \boldsymbol{u}_k^{\mathsf{H}} \boldsymbol{v}_{ji}|^2 + \|\boldsymbol{u}_k\|_2^2 \sigma^2}.$$

Based on the singular value decomposition (SVD) $\boldsymbol{X}^* = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}$, the transmit beamformer is simply chosen as $\boldsymbol{U}\boldsymbol{\Sigma}^{\frac{1}{2}}$, and receive beamformer is given by $\boldsymbol{V}\boldsymbol{\Sigma}^{\frac{1}{2}}$ with power normalization. Fig. 7b shows the achievable sum-rate over different transmit power P. Each point is averaged across 100 channel realizations with random $\mathcal{E}$ ($p = 0.2$). The result also demonstrates that the proposed RTR algorithm is capable of achieving high data rates by leveraging transmitter cooperation.

In summary, our numerical experiments demonstrate that the proposed Riemannian trust-region algorithm is capable of obtaining high-precision solutions with second-order stationary points, leading to high achievable DoFs and data rates. Furthermore, the computation time of RTR algorithm in medium scale problems is comparable with first order algorithms, which demonstrates that the proposed RTR algorithm is a powerful algorithm capable of harnessing the benefit of topological cooperation in problems involving medium network sizes.

VI. Conclusions

This work investigates the opportunities of transmitter cooperation based only on topological information with message sharing. Our contributions include the derivation of a generalized topological interference alignment condition, followed by the development of a low-rank matrix optimization approach to maximize the achievable DoFs. To solve the resulting generalized low-rank optimization problem which is nonconvex in complex field, we developed Riemannian optimization algorithms by exploiting the complex non-compact Stiefel manifold for fixed-rank matrices in complex field. In particular, we adopted the semidefinite lifting technique and Burer-Monteiro factorization approach. Our experiments demonstrated that the proposed Riemannian algorithms considerably outperformed the alternating minimization algorithm. Additionally, the proposed Riemannian trust-region algorithm achieves high DoFs with high-precision second-order stationary point solutions, with computation complexity comparable with the first-order Riemannian conjugate gradient algorithm.

Appendix A
Proof of Proposition 2

For simplicity, we only give the proof of the single data stream case, while it can be readily extended to general multiple data stream cases. In this case, $\boldsymbol{X}$ is given by

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{u}_1^{\mathsf{H}}\boldsymbol{v}_{11} & \cdots & \boldsymbol{u}_1^{\mathsf{H}}\boldsymbol{v}_{KK} \\ \vdots & \ddots & \vdots \\ \boldsymbol{u}_K^{\mathsf{H}}\boldsymbol{v}_{11} & \cdots & \boldsymbol{u}_K^{\mathsf{H}}\boldsymbol{v}_{KK} \end{bmatrix} \in \mathbb{C}^{K\times K^2}. \tag{65}$$

Let $\boldsymbol{x}_i$ denote the transpose of the i-th row of matrix $\boldsymbol{X}$. Problem (19) can be rewritten as

$$\begin{aligned} &\underset{\boldsymbol{X}\in\mathbb{C}^{m\times n}}{\text{minimize}} \ \|\boldsymbol{X}\|_* \\ &\text{subject to } \mathbf{1}^{\mathsf{H}}\boldsymbol{x}_k^{\mathcal{D}_k} = 1, \quad \forall k \in [K] \\ &\qquad\quad \boldsymbol{x}_k^{\mathcal{G}_k} = \mathbf{0}, \quad \forall k \in [K], \end{aligned} \tag{66}$$

in which $\boldsymbol{x}_k^{\mathcal{D}_k}, \boldsymbol{x}_k^{\mathcal{G}_k}$ are vectors whose elements are sampled from $\boldsymbol{x}_k$, and $\mathcal{D}_k, \mathcal{G}_k$ are the index sets of sampling. $\mathcal{D}_k$ and $\mathcal{G}_k$ are given by

$$\mathcal{D}_k = \{(j-1)*K + k : (k,j) \in \mathcal{E}, k \in \mathcal{S}_j\} \tag{67}$$

and

$$\mathcal{G}_k = \{(j-1)*K + i : i \neq k, (k,j) \in \mathcal{E}, i \in \mathcal{S}_j\}, \tag{68}$$

respectively.

Lemma 1: The optimal solution of (66), denoted by $\boldsymbol{X}^\star$, is given by

$$\boldsymbol{x}^{\star}{}_k^{\mathcal{D}_k} = \frac{1}{|\mathcal{D}_k|}\mathbf{1}, \tag{69}$$

where $\mathbf{1}$ denotes the vector of all ones. The remaining entries of $\boldsymbol{X}^\star$ are all zeros.

Proof: Let $g(\boldsymbol{X}) = \|\boldsymbol{X}\|_* = \mathrm{Tr}(\sqrt{\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}})$. To proof Lemma 1, it is equivalent to prove that $t = 0$ is a minimum of the convex function

$$h(t) = \|\boldsymbol{X}^\star + t\boldsymbol{X}\|_*, \quad \mathcal{A}(\boldsymbol{X}^\star + t\boldsymbol{X}) = \boldsymbol{b}, \ t \in \mathbb{R}. \tag{70}$$

This can be deduced from the fact that any feasible point can be expressed as $\boldsymbol{X}^\star + t\boldsymbol{X}$, and if $t = 0$ is a minimum of $h(t)$, then $g(\boldsymbol{X}^\star) \leq g(\boldsymbol{X})$ always holds.

Based on the structure of $\mathcal{D}_k$ and $\mathcal{G}_k$, we have

$$\boldsymbol{X}^\star\boldsymbol{X}^{\star\mathsf{H}} = \begin{bmatrix} \frac{1}{|\mathcal{D}_1|} & 0 & \cdots & 0 \\ 0 & \frac{1}{|\mathcal{D}_2|} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{|\mathcal{D}_K|} \end{bmatrix}. \tag{71}$$

Then $\boldsymbol{X}^\star$ is a full rank matrix, and we can find $|t| \leq \epsilon$ such that $\boldsymbol{X}^\star + t\boldsymbol{X}$ is invertible. Therefore, the derivative of h is given by

$$\begin{aligned} h'(t) = \frac{1}{2}\big\langle \left((\boldsymbol{X}^\star + t\boldsymbol{X})(\boldsymbol{X}^\star + t\boldsymbol{X})^{\mathsf{H}}\right)^{-\frac{1}{2}}, \\ \boldsymbol{X}\boldsymbol{X}^{\star\mathsf{H}} + \boldsymbol{X}^\star\boldsymbol{X}^{\mathsf{H}} + 2t\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}\big\rangle. \end{aligned} \tag{72}$$

Then we have

$$h'(0) = \frac{1}{2}\big\langle \left(\boldsymbol{X}^\star\boldsymbol{X}^{\star\mathsf{H}}\right)^{-\frac{1}{2}}, \boldsymbol{X}\boldsymbol{X}^{\star\mathsf{H}} + \boldsymbol{X}^\star\boldsymbol{X}^{\mathsf{H}}\big\rangle. \tag{73}$$

Since $\mathcal{A}(\boldsymbol{X}^\star + t\boldsymbol{X}) = \boldsymbol{b}$, i.e.,

$$\mathbf{1}^{\mathsf{H}}(\boldsymbol{x}^{\star}{}_k^{\mathcal{D}_k} + t\boldsymbol{x}_k^{\mathcal{D}_k}) = 1 \tag{74}$$

then we have $\langle \boldsymbol{x}^{\star}{}_k^{\mathcal{D}_k}, \boldsymbol{x}_k^{\mathcal{D}_k}\rangle = 0$. Therefore,

$$\mathrm{diag}(\boldsymbol{X}\boldsymbol{X}^{\star\mathsf{H}}) = \mathrm{diag}(\boldsymbol{X}^\star\boldsymbol{X}^{\mathsf{H}}) = \mathbf{0}. \tag{75}$$

From (75) (71), we can deduce that $h'(0) = 0$, and thus $t = 0$ is a minimum of $h(t)$. □

From Lemma 1 we know that the optimal solution $\boldsymbol{X}^\star$ of (19) is full rank. So nuclear norm relaxation approach always fails.

Appendix B
Proof of Proposition 4

The elements in the vertical space $\mathcal{V}_{\boldsymbol{Y}}$ must be tangential to the equivalent class $[\boldsymbol{Y}] = \{\boldsymbol{Y}\boldsymbol{Q} : \boldsymbol{Q}^{\mathsf{H}}\boldsymbol{Q} = \boldsymbol{I}\}$. Let $\boldsymbol{Y}(t) = \boldsymbol{Y}_0\boldsymbol{Q}(t)$ be a curve in $[\boldsymbol{Y}_0]$ through $\boldsymbol{Y}_0$ at $t = 0$, i.e., $\boldsymbol{Q}(0) = \boldsymbol{I}$. Then we have

$$\boldsymbol{Y}(t)\boldsymbol{Y}(t)^{\mathsf{H}} = \boldsymbol{Y}_0\boldsymbol{Q}(t)\boldsymbol{Q}(t)^{\mathsf{H}}\boldsymbol{Y}_0^{\mathsf{H}} = \boldsymbol{Y}_0\boldsymbol{Y}_0^{\mathsf{H}}. \tag{76}$$

By differentiating (76) we get

$$\dot{\boldsymbol{Y}}(t)\boldsymbol{Y}(t)^{\mathsf{H}} + \boldsymbol{Y}(t)\dot{\boldsymbol{Y}}(t)^{\mathsf{H}} = \mathbf{0}. \tag{77}$$

So we deduce that $\dot{\boldsymbol{Y}}(0)$ belongs to

$$\{\boldsymbol{Z} \in \mathbb{C}^{N\times r} : \boldsymbol{Z}\boldsymbol{Y}_0^{\mathsf{H}} + \boldsymbol{Y}_0\boldsymbol{Z}^{\mathsf{H}} = \boldsymbol{0}\}, \tag{78}$$

of which $\mathcal{T}_{\boldsymbol{Y}}\mathcal{M}$ is a subset. On the other side, let $F : \boldsymbol{Y} \mapsto \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$, then (78) is $\ker(\mathrm{D}F(\boldsymbol{Y_0}))$ and $F^{-1}(\boldsymbol{Y}_0\boldsymbol{Y}_0^{\mathsf{H}}) = [\boldsymbol{Y}_0]$. Therefore, from [23, Sec. 3.5.7] we know

$$\mathcal{T}_{\boldsymbol{Y}}\mathcal{M} = \{\boldsymbol{Z} \in \mathbb{C}^{N\times r} : \boldsymbol{Z}\boldsymbol{Y}_0^{\mathsf{H}} + \boldsymbol{Y}_0\boldsymbol{Z}^{\mathsf{H}} = \boldsymbol{0}\}. \tag{79}$$

Without loss of generality, we can set

$$\dot{\boldsymbol{Y}}(t) = \boldsymbol{Y}(t)\boldsymbol{\Omega}(t), \quad \boldsymbol{\Omega}(t) \in \mathbb{C}^{r\times r}, \tag{80}$$

since $\boldsymbol{Y}(t) \in \mathbb{C}_*^{N\times r}$ is full rank. Then we can replace equation (80) in (77) and obtain

$$\boldsymbol{Y}(t)(\boldsymbol{\Omega}(t) + \boldsymbol{\Omega}(t)^{\mathsf{H}})\boldsymbol{Y}(t)^{\mathsf{H}} = \boldsymbol{0}. \tag{81}$$

Therefore, the vertical space is given by

$$\mathcal{V}_{\boldsymbol{Y}} = \{\boldsymbol{Y}_0\boldsymbol{\Omega} : \boldsymbol{\Omega}^{\mathsf{H}} = -\boldsymbol{\Omega}\}. \tag{82}$$

APPENDIX C
PROOF OF PROPOSITION 5

The horizontal space is given by

$$\mathcal{H}_{\boldsymbol{Y}} = \{\overline{\boldsymbol{\xi}} \in \mathcal{T}_{\boldsymbol{Y}}\overline{\mathcal{M}} : \overline{g}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}}, \overline{\boldsymbol{\zeta}}) = 0,\ \forall \overline{\boldsymbol{\zeta}} \in \mathcal{V}_{\boldsymbol{Y}}\}, \tag{83}$$

that is

$$\overline{g}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}}, \boldsymbol{Y}\boldsymbol{\Omega}) = 0, \quad \forall \boldsymbol{\Omega}^{\mathsf{H}} = -\boldsymbol{\Omega}. \tag{84}$$

Since

$$\begin{aligned}\overline{g}_{\boldsymbol{Y}}(\overline{\boldsymbol{\xi}}, \boldsymbol{Y}\boldsymbol{\Omega}) &= \mathrm{Tr}(\overline{\boldsymbol{\xi}}^{\mathsf{H}}\boldsymbol{Y}\boldsymbol{\Omega} + \boldsymbol{\Omega}^{\mathsf{H}}\boldsymbol{Y}^{\mathsf{H}}\overline{\boldsymbol{\xi}}) \\ &= \mathrm{Tr}(\overline{\boldsymbol{\xi}}^{\mathsf{H}}\boldsymbol{Y}\boldsymbol{\Omega} - \boldsymbol{\Omega}\boldsymbol{Y}^{\mathsf{H}}\overline{\boldsymbol{\xi}}) \\ &= \mathrm{Tr}((\overline{\boldsymbol{\xi}}^{\mathsf{H}}\boldsymbol{Y} - \boldsymbol{Y}^{\mathsf{H}}\overline{\boldsymbol{\xi}})\boldsymbol{\Omega}),\end{aligned} \tag{85}$$

we know the horizontal space consists of all the elements $\overline{\boldsymbol{\xi}}$ that satisfies $\mathrm{Tr}((\overline{\boldsymbol{\xi}}^{\mathsf{H}}\boldsymbol{Y} - \boldsymbol{Y}^{\mathsf{H}}\overline{\boldsymbol{\xi}})\boldsymbol{\Omega}) = 0$ for all $\boldsymbol{\Omega}^{\mathsf{H}} = -\boldsymbol{\Omega}$. Therefore, the horizontal space is

$$\mathcal{H}_{\boldsymbol{Y}} = \{\boldsymbol{\xi} \in \mathbb{C}^{N\times r} : \boldsymbol{\xi}^{\mathsf{H}}\boldsymbol{Y} = \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\xi}\}. \tag{86}$$

Suppose for a vector $\boldsymbol{\xi} \in \mathcal{T}_{\boldsymbol{Y}}\mathcal{M}$ its projection onto the vertical space is given by $\boldsymbol{\xi}^v = \boldsymbol{Y}\boldsymbol{\Omega}_{\boldsymbol{\xi}}$, then the horizontal projection is given by $\boldsymbol{\xi}^h = \boldsymbol{\xi} - \boldsymbol{Y}\boldsymbol{\Omega}_{\boldsymbol{\xi}}$, and

$${\boldsymbol{\xi}^h}^{\mathsf{H}}\boldsymbol{Y} = \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\xi}^h. \tag{87}$$

So we can find the $\boldsymbol{\Omega}_{\boldsymbol{\xi}}$ from

$$\begin{aligned}(\boldsymbol{\xi} - \boldsymbol{Y}\boldsymbol{\Omega}_{\boldsymbol{\xi}})^{\mathsf{H}}\boldsymbol{Y} &= \boldsymbol{Y}^{\mathsf{H}}(\boldsymbol{\xi} - \boldsymbol{Y}\boldsymbol{\Omega}_{\boldsymbol{\xi}}) \\ \Rightarrow \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Y}\boldsymbol{\Omega}_{\boldsymbol{\xi}} + \boldsymbol{\Omega}_{\boldsymbol{\xi}}\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Y} &= \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\xi} - \boldsymbol{\xi}^{\mathsf{H}}\boldsymbol{Y}.\end{aligned} \tag{88}$$

Then we conclude that the horizontal projection of $\boldsymbol{\xi}$ is given by

$$\Pi_{\boldsymbol{Y}}^h\boldsymbol{\xi} = \boldsymbol{\xi} - \boldsymbol{Y}\boldsymbol{\Omega}, \tag{89}$$

where $\boldsymbol{\Omega}$ is the solution to the Lyapunov equation

$$\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Y}\boldsymbol{\Omega} + \boldsymbol{\Omega}\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{Y} = \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\xi} - \boldsymbol{\xi}^{\mathsf{H}}\boldsymbol{Y}. \tag{90}$$

APPENDIX D
COMPUTING THE RIEMANNIAN GRADIENT AND HESSIAN

We first rewrite the objective function of (29) as

$$f(\boldsymbol{Y}) = \frac{1}{2}\sum_{i=1}^{l} |\langle \boldsymbol{B}_i, \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\rangle - b_i|^2. \tag{91}$$

The complex gradient of $f(\boldsymbol{Y})$ is given by

$$\begin{aligned}f'(\boldsymbol{Y}) &= \sum_{i=1}^{l}(\langle \boldsymbol{B}_i, \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\rangle - b_i)\boldsymbol{B}_i\boldsymbol{Y} \\ &\quad + (\langle \boldsymbol{B}_i^{\mathsf{H}}, \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\rangle - b_i^*)\boldsymbol{B}_i^{\mathsf{H}}\boldsymbol{Y} \\ &= \sum_{i=1}^{l}(C_i\boldsymbol{B}_i + C_i^*\boldsymbol{B}_i^{\mathsf{H}})\boldsymbol{Y},\end{aligned} \tag{92}$$

in which $C_i = \langle \boldsymbol{B}_i, \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\rangle - b_i$. The Riemannian gradient $\overline{\mathrm{grad}} f(\boldsymbol{Y})$ is derived from (44), and we find that

$$\begin{aligned}\mathrm{D}f(\boldsymbol{Y})[\boldsymbol{\xi}] &= \frac{1}{2}\sum_{i=1}^{l}\langle \boldsymbol{B}_i, \boldsymbol{\xi}\boldsymbol{Y}^{\mathsf{H}} + \boldsymbol{Y}\boldsymbol{\xi}^{\mathsf{H}}\rangle^*(\langle \boldsymbol{B}_i, \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\rangle - b_i) \\ &\quad + (\langle \boldsymbol{B}_i, \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\rangle - b_i)^*\langle \boldsymbol{B}_i, \boldsymbol{\xi}\boldsymbol{Y}^{\mathsf{H}} + \boldsymbol{Y}\boldsymbol{\xi}^{\mathsf{H}}\rangle \\ &= \overline{g}_{\boldsymbol{Y}}((C_i\boldsymbol{B}_i + C_i^*\boldsymbol{B}_i^{\mathsf{H}})\boldsymbol{Y}, \boldsymbol{\xi}).\end{aligned} \tag{93}$$

Therefore, $\overline{\mathrm{grad}} f(\boldsymbol{Y}) = f'(\boldsymbol{Y})$. Then we observe that $\overline{\mathrm{grad}} f(\boldsymbol{Y})^{\mathsf{H}}\boldsymbol{Y} = \boldsymbol{Y}^{\mathsf{H}}\overline{\mathrm{grad}} f(\boldsymbol{Y})$, i.e., $\overline{\mathrm{grad}} f(\boldsymbol{Y})$ is already in the horizontal space $\mathcal{V}_{\boldsymbol{Y}}$. So the horizontal representation of Riemannian gradient is given by

$$\mathrm{grad} f(\boldsymbol{Y}) = \sum_{i=1}^{l}(C_i\boldsymbol{B}_i + C_i^*\boldsymbol{B}_i^{\mathsf{H}})\boldsymbol{Y}. \tag{94}$$

To derive the Riemannian Hessian (47), we compute

$$\begin{aligned}\mathrm{D}\overline{\mathrm{grad}} f(\boldsymbol{Y})[\boldsymbol{\eta}_{\boldsymbol{Y}}] &= \sum_{i=1}^{l}(C_{\boldsymbol{\eta}_i}\boldsymbol{B}_i\boldsymbol{Y} + C_i\boldsymbol{B}_i\boldsymbol{\eta}_Y \\ &\quad + C_{\boldsymbol{\eta}_i}^*\boldsymbol{B}_i^{\mathsf{H}}\boldsymbol{Y} + C_i^*\boldsymbol{B}_i^{\mathsf{H}}\boldsymbol{\eta}_{\boldsymbol{Y}}),\end{aligned} \tag{95}$$

where $C_{\boldsymbol{\eta}_i} = \langle \boldsymbol{B}_i, \boldsymbol{Y}\boldsymbol{\eta}_{\boldsymbol{Y}}^{\mathsf{H}} + \boldsymbol{\eta}_{\boldsymbol{Y}}\boldsymbol{Y}^{\mathsf{H}}\rangle$. We conclude that

$$\begin{aligned}\mathrm{Hess} f(\boldsymbol{Y})[\boldsymbol{\eta}_{\boldsymbol{Y}}] = \Pi_{\boldsymbol{Y}}^h\Bigg(&\sum_{i=1}^{l}(C_{\boldsymbol{\eta}_i}\boldsymbol{B}_i\boldsymbol{Y} + C_i\boldsymbol{B}_i\boldsymbol{\eta}_Y \\ &+ C_{\boldsymbol{\eta}_i}^*\boldsymbol{B}_i^{\mathsf{H}}\boldsymbol{Y} + C_i^*\boldsymbol{B}_i^{\mathsf{H}}\boldsymbol{\eta}_{\boldsymbol{Y}})\Bigg).\end{aligned} \tag{96}$$

REFERENCES

[1] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile Internet," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

[2] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-scale convex optimization for dense wireless cooperative networks," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4729–4743, Sep. 2015.

[3] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

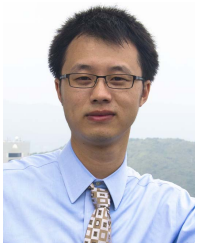
[4] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[5] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green Cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

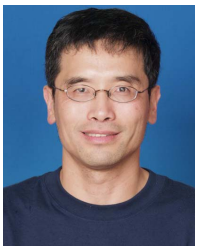
[6] K. Yang, Y. Shi, and Z. Ding, "Generalized matrix completion for low complexity transceiver processing in cache-aided Fog-RAN via the Burer-Monteiro approach," *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 863–867.

[7] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.
[8] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, Jul. 2012.
[9] Y. Shi, J. Zhang, and K. B. Letaief, "Optimal stochastic coordinated beamforming for wireless cooperative networks with CSI uncertainty," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 960–973, Feb. 2015.
[10] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," *Math. Program.*, vol. 157, no. 2, pp. 515–545, Jun. 2016.
[11] A. G. Davoodi and S. A. Jafar, "Generalized degrees of freedom of the symmetric K user interference channel under finite precision CSIT," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6561–6572, Oct. 2017.
[12] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
[13] S. A. Jafar, "Topological interference management through index coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 529–568, Jan. 2014.
[14] X. Yi and D. Gesbert, "Topological interference management with transmitter cooperation," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6107–6130, Nov. 2015.
[15] Y. Shi, J. Zhang, and K. B. Letaief, "Low-rank matrix completion for topological interference management by Riemannian pursuit," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4703–4717, Jul. 2016.
[16] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 608–622, Jun. 2016.
[17] G. Sridharan and W. Yu, "Linear beamformer design for interference alignment via rank minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5910–5923, Nov. 2015.
[18] D. S. Papailiopoulos and A. G. Dimakis, "Interference alignment as a rank constrained rank minimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4278–4288, Aug. 2012.
[19] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, p. 717, Dec. 2009.
[20] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proc. 45th Annu. ACM Symp. Theory Comput.*, Jun. 2013, pp. 665–674.
[21] R. Ge, J. D. Lee, and T. Ma, "Matrix completion has no spurious local minimum," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 2973–2981.
[22] X. Yi and G. Caire, "Topological coded caching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2039–2043.
[23] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[24] N. Boumal, P.-A. Absil, and C. Cartis, "Global rates of convergence for nonconvex optimization on manifolds," *IMA J. Numer. Anal.*, vol. 39, no. 1, pp. 1–33, Feb. 2018.
[25] R. Ge, C. Jin, and Y. Zheng, "No spurious local minima in nonconvex low rank problems: A unified geometric analysis," *Proc. 34th Int. Conf. Mach. Learn.-Volume*, Aug. 2018, pp. 1233–1242.
[26] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Math. Program.*, vol. 95, no. 2, pp. 329–357, Feb. 2003.
[27] G. Bresler, D. Cartwright, and D. Tse, "Feasibility of interference alignment for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5573–5586, Sep. 2014.
[28] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 812–821, Feb. 2012.
[29] B. Hassibi, "Topological interference alignment in wireless networks," in *Proc. Smart Antennas Workshop*, Aug. 2014, pp. 1–143.
[30] Y. Zhang and Z. Lu, "Penalty decomposition methods for rank minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, vol. 30, no. 3, pp. 46–54.
[31] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[32] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *J. Optim. Theory Appl.*, vol. 169, no. 3, pp. 1042–1068, Jun. 2016.
[33] R. Bamler and S. Mandt, "Improving optimization for models with continuous symmetry breaking," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 80, Jul. 2018, pp. 423–432.
[34] Y. Shi, B. Mishra, and W. Chen, "Topological interference management with user admission control via Riemannian optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7362–7375, Nov. 2017.
[35] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre, "Low-rank optimization on the cone of positive semidefinite matrices," *SIAM J. Optim.*, vol. 20, no. 5, pp. 2327–2351, May 2010.
[36] S. Yatawatta, "Radio interferometric calibration using a Riemannian manifold," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 3866–3870.
[37] W. Barth, K. Hulek, C. Peters, and A. Van de Ven, "Compact complex surfaces," in *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge* (A Series of Modern Surveys in Mathematics), vol. 4. Springer, 2015.
[38] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Nat. Bur. Standards*, vol. 49, no. 6, pp. 409–436, Dec. 1952.
[39] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a MATLAB toolbox for optimization on manifolds," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1455–1459, 2014.

Kai Yang (S'16) received the B.S. degree in electronic engineering from the Dalian University of Technology, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Information Science and Technology, ShanghaiTech University. His research interests are in big data processing, mobile edge/fog computing, mobile edge artificial intelligence, and dense communication networking.



Yuanming Shi (S'13–M'15) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2011, and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology (HKUST) in 2015. Since 2015, he has been with the School of Information Science and Technology, ShanghaiTech University, where he is currently a tenured Associate Professor. He visited the University of California at Berkeley, Berkeley, CA, USA, from 2016 to 2017. His research areas include optimization, statistics, machine learning, signal processing, and their applications to wireless communications and quantitative finance. He was a recipient of the 2016 IEEE Marconi Prize Paper Award in Wireless Communications and the 2016 Young Author Best Paper Award by the IEEE Signal Processing Society.



Zhi Ding (S'88–M'90–SM'95–F'03) received the Ph.D. degree in electrical engineering from Cornell University in 1990. From 1990 to 2000, he was a Faculty Member with Auburn University and then with the University of Iowa. He held visiting positions at Australian National University, The Hong Kong University of Science and Technology, the NASA Lewis Research Center, and the USAF Wright Laboratory. He has active collaboration with researchers from Australia, Canada, China, Finland, Hong Kong, Japan, South Korea, Singapore, and Taiwan. He is currently a Professor of electrical and computer engineering with the University of California at Davis, Davis, CA, USA. He is the co-author of the text book *Modern Digital and Analog Communication Systems* (Oxford University Press, 5th edition, 2019). He was a member of the Technical Committee on Statistical Signal and Array Processing and the Technical Committee on Signal Processing for Communications from 1994 to 2003. He has been an Active Member of the IEEE, serving on the technical program committees of several workshops and conferences. He was the Technical Program Chair of the 2006 IEEE Globecom and the General Chair of the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1994 to 1997 and from 2001 to 2004 and the IEEE SIGNAL PROCESSING LETTERS from 2002 to 2005. He was also an IEEE Distinguished Lecturer of the Circuits and Systems Society from 2004 to 2006 and the Communications Society from 2008 to 2009. He has served as the IEEE Transactions on Wireless Communications Steering Committee Member from 2007 to 2009 and as its Chair from 2009 to 2010.