Interference Alignment for One-hop and Two-hops MIMO Systems with Uncoordinated Interference

Siavash Mollaebrahim, Pouya M. Ghari, Mohammad Sadegh Fazel, Glauber Brante, *Member, IEEE*, Muhammad Ali Imran, *Senior Member, IEEE*

***Abstract*—Providing higher data rate is a momentous goal for wireless communications systems, while interference is an important obstacle to reach this purpose. To cope with this problem, interference alignment (IA) has been proposed. In this paper, we propose two rank minimization methods to enhance the performance of IA in the presence of uncoordinated interference, *i.e.*, interference that cannot be properly aligned with the rest of the network and thus is a crucial issue. In this scenario, we consider perfect and imperfect channel state information (CSI) cases. Our proposed approaches employ the l_2 and the Schatten-p norms to approximate the rank function, due to its non-convexity. We also use a new convex relaxation to expand the feasible set of our optimization problem, providing lower rank solutions compared to other IA methods from the literature. In addition, we propose a modified weighted-sum method to deal with interference in the relay-aided MIMO interference channel, which employs a set of weighting parameters in order to find more solutions.**

***Index Terms*—Interference alignment, MIMO interference channel, relay-aided MIMO interference channel, convex relaxation.**

I. Introduction

INTERFERENCE management is a crucial issue in wireless networks, especially in heterogeneous environments consisting of smaller cells deployed to meet the explosively increasing traffic demand, where interference poses a serious obstacle for wireless communications [1]. Then, several methods have been proposed in the literature to deal with interference, which are either based on accounting other users' signals as noise, or trying to orthogonalize the communication links. Using any of these two approaches in a K-user MIMO system gives each user a portion of $1/K$ of all available resources. Consequently, when the number of users increases, the amount of resources devoted to each user considerably decreases [2].

On the other hand, interference alignment (IA) is a promising technique that exploits the spatial dimensions offered by multiple antennas to coordinate transmissions, such that the interference at each receiver can be reduced. A measure of the sum-capacity of interference channels is given by the degrees of freedom (DoF), also denoted as multiplexing gain in [3], which is defined as $d = \lim_{SNR \to \infty} \frac{C_{\mathrm{sum}}(\mathrm{SNR})}{\log(\mathrm{SNR})}$ where $C_{\mathrm{sum}}(\mathrm{SNR})$ is the sum capacity of the network. Also, the DoF can be interpreted as the number of interference-free signaling dimensions [4]. For instance, in a K-user MIMO system where each user has N antennas, it has been shown that IA can yield $KN/2$ DoF [5]. However, this result is highly dependent on the channel extensions, which is not usually available, while it can also increase the implementation complexity [6]. The main idea behind IA is to design precoding and receiver filter matrices in order to minimize the dimension of the interfering signals. In addition, in contrast to other methods that require the decomposition of multi-antenna nodes and infinite symbol extensions, as in [3], linear beamforming-based interference alignment schemes are simpler to implement [7].

Recently, various methods have been proposed to design linear precoding and receive matrices for IA, as in [4], [7]–[11]. For instance, the authors in [4] propose a rank constrained rank minimization (RCRM) approach, whose goal is to minimize the rank of the interference matrix, which consequently reduces the dimension occupied by the interference. In addition, [4] also shows that the RCRM approach is equivalent to maximize the DoF in MIMO interference channels. Nevertheless, due to the non-convexity and NP-hardness of the RCRM problem, the nuclear norm has been used in [4] as an approximation for the rank function, which was replaced in [8] by the sum of log functions. Alternatively, other approaches have employed algorithms to optimize objective functions, *e.g.*, SINR maximization [9], or interference leakage minimization [10].

An important issue for IA techniques is the uncoordinated interference, which severely degrades the performance of the coordinated part of the network [12]. For instance, in heterogeneous pico-cell networks, the interference caused by femto and home base stations are usually sources of uncoordinated interference, whose associated users are not always able to cooperate. Thus, their interference cannot be fully aligned.

Furthermore, interference is also an important issue for relay-aided MIMO networks, since it impacts the signal received at the relay and at the destination simultaneously. Moreover, the design of relay processing matrices becomes a complicated task [13]. Unfortunately, previous single-hop IA algorithms are not easily applicable for relay-aided scenarios, once the power constraints of relays depend on the precoding matrices at transmitters, as well as on the processing

S. Mollaebrahim, M. Fazel are with the Department of Electrical and Computer Engineering, Isfahan University of Technology (IUT), Isfahan, Iran, e-mail: siavashmol@gmail.com, fazel@cc.iut.ac.ir.

P. M. Ghari is with the Electrical and Computer Engineering Department, University of Tehran (UT), Tehran, Iran, e-mail: p.m.ghari@ut.ac.ir.

G. Brante is with Federal University of Technology-Paraná, Curitiba, Brazil, e-mail: gbrante@utfpr.edu.br.

M. A. Imran is with the School of Engineering, University of Glasgow, UK, e-mail: Muhammad.Imran@glasgow.ac.uk.

This research was partly funded by EPSRC Global Challenge Research Fund-the Dare project-EP/P028764/1

matrices at the relays. Recently, the number of methods has been proposed to design processing matrices for relay-aided interference channels such as [6], [14]–[16]. For instance, in [14] the authors propose two IA approaches for an amplify-and-forward (AF) relay channel, extending previous proposals for the single-hop scenario [10], [17]. AF relaying has been chosen due to its lower complexity and shorter delay compared to decode-and-forward (DF) relaying [18]. Then, in the first approach the sum power of the interference at the receiver is minimized, while the second approach aims at minimizing the weighted-sum mean square errors (WMSE).

A. Contributions

In this paper, we design IA techniques in scenarios subjected to uncoordinated sources of interference. Then, two examples of these scenarios would be the heterogeneous networks, which can be modeled as a MIMO interference channel where interference of neighbor femto-cells are not coordinated with interference of the macro BS; and a relay-aided scenario, where the noise enhanced by relays can also be seen as uncoordinated interference.

The main contribution of this paper relates to three issues. Firstly, the rank function of a semidefinite matrix is defined by the number of non-zero singular values of it, and because it is non-convex, we need to use a surrogate to minimize it. The surrogate of the rank function defines how we treat the singular values. The first contribution of this paper relates to this issue. We use the l_2 norm of singular values of the interference matrix to minimize the rank of it, in which larger singular values will have higher weights than in the l_1 norm [4]. This difference in relative weightings for small and large residuals is reflected in the solutions of the associated approximation problems. For instance, when the effect of interference is strong the interference matrix can have larger singular values, and thus minimizing the l_2 norm is a reasonable choice because it penalizes the large singular values more than the smaller ones. Moreover, in the second method we use the Schatten-p norm as the surrogate of the rank function. The Schatten-p norm is a proper candidate because its behavior is more similar to the rank function (it treats the singular values comparatively equally) than the nuclear norm [4] and the log function [8]. Thus, using the Schatten-p norm can help us to obtain lower rank solutions which means that the dimension occupied by the interference matrix becomes smaller, and higher DoF interpreted as the number of interference-free signaling dimensions can be obtained.

Secondly, in the RCRM framework, the signal matrix should be full rank. Authors in [4], [8] have restricted their analysis to positive definite matrices which are full rank. However, there are matrices which are not positive definite but still they are full rank. Considering those matrices enables us to expand the feasible set of our optimization problem. Expanding the feasible set of an optimization problem can lead to improved results, once more candidates to the solution are considered.

Thus, we propose a new convex constraint that yields two important improvements in the IA context. First, the proposed constraint expands the feasible set of our optimization problem compared to the other approaches in the literature (*e.g.*, [4]), which allows increased performance in terms of the sum-rate at medium and high SNR. Second, the proposed constraint provides lower rank solutions with respect to the interference matrix, consequently obtaining better DoF. We also propose a new method to deal with interference in the AF relay-aided MIMO system, for which a modified weighted-sum method is proposed. The modified approach employs a set of weighting parameters in order to find more solutions which can increase our abality to get better results.

Thirdly, in comparison with [4], [8]–[10], we consider the effect of imperfect CSI for coordinated and uncoordinated cases, and a method has been proposed to deal with it. In fact, since the performance of IA methods degrades in the present of imperfect CSI, proposing approaches to deal with it is important. In this case, we face enhanced noise whose power scales with the desired signal power which degrades the performance of IA.

Notation: The expectation operator is denoted by $E[.]$. The Hermitian transpose of a matrix $\boldsymbol{A}$, its trace and its i-th largest singular value are denoted by $\boldsymbol{A}^H$, $\mathrm{tr}(\boldsymbol{A})$ and $\delta_i(\boldsymbol{A})$, respectively. Moreover, $\mathbf{1}_a$ and $\boldsymbol{I}_a$ are the square matrix of ones and the identity of order a, respectively, while $\mathbf{0}_{a\times b}$ is an $a \times b$ matrix of zeros. Finally, $\boldsymbol{A}(i,j)$ denotes the element in the i-th row and the j-th column of $\boldsymbol{A}$.

II. Proposed Methods for MIMO Interference Channels

A. System Model for MIMO Interference Channel

Let us consider a K-user interference channel with K transmit-receive pairs and M sources of uncoordinated interference. As we stated, in heterogeneous networks, femtocells and home base stations may cause uncoordinated interference. For instance, in the LTE heterogeneous network, traditional base stations and picocells coexist with femtocells and home base stations. In this network, there are interfaces S1 and S11 which support data traffic between the corresponding nodes that create uncoordinated interference. By using those interfaces, information of uncoordinated interference can be obtained [12]. Each transmitter and receiver has N_t and N_r antennas, respectively. Moreover, transmitters are assumed to be synchronized and each user transmits the symbol vector $\boldsymbol{s}_k \in \mathbb{R}^{d\times 1}$, $k = 1, \cdots, K$, to its associated receiver. Indeed, IA needs synchronization to avoid any timing and carrier frequency offsets. In the absence of synchronization, additional interference terms are added to our system, and make it too complicated. To meet this requirement, some synchronization strategies such as using GPS satellite signals or software defined radio (SDR) implementation have been introduced [19]–[21]. Also, $(N_t \times N_r, d)^K$ denotes this K-user MIMO system, where d is the DoF intended by each user [22], *i.e.*, the number of signal space dimensions that are free of interference which is denoted as multiplexing gain [4]. Then, the received signal at the k-th receiver after linear processing is

$$\boldsymbol{y}_k = \boldsymbol{W}_k^H \boldsymbol{H}_{k,k} \boldsymbol{F}_k \boldsymbol{s}_k + \boldsymbol{W}_k^H \sum_{l=1,l\neq k}^{K} \boldsymbol{H}_{k,l} \boldsymbol{F}_l \boldsymbol{s}_l + \boldsymbol{W}_k^H \sum_{m=1}^{M} \widetilde{\boldsymbol{H}}_{k,m} \widetilde{\boldsymbol{F}}_m \ \boldsymbol{x}_m + \boldsymbol{W}_k^H \boldsymbol{n}_k \quad (1)$$

where $\boldsymbol{H}_{k,l} \in \mathbb{R}^{N_r \times N_t}$ represents the channel between the k-th receiver and the l-th transmitter, while $\widetilde{\boldsymbol{H}}_{k,m} \in \mathbb{R}^{N_r \times N_m}$ is the channel between the k-th receiver and the m-th uncoordinated source. Moreover, $\boldsymbol{W}_k \in \mathbb{R}^{N_r \times d}$ is the linear receive filter, $\boldsymbol{F}_k \in \mathbb{R}^{N_t \times d}$ and $\widetilde{\boldsymbol{F}}_m \in \mathbb{R}^{N_m \times d_m}$ are the precoding matrices at the k-th transmitter and at the m-th uncoordinated source, respectively, $\boldsymbol{x}_m \in \mathbb{R}^{d_m \times 1}$ denotes the symbol vector of the m-th uncoordinated source, and $\boldsymbol{n}_k$ is the zero-mean complex additive white Gaussian noise with covariance $\sigma_k^2 \boldsymbol{I}_{N_r}$.

Generally, IA relies on channel state information (CSI), and most IA approaches have assumed of having perfect CSI knowledge at the transmitter and receiver nodes. However, this assumption is unpractical due to the channel estimation errors, feedback delay, etc. Thus, in this paper we consider the effect of imperfect CSI caused by channel estimation errors.

To do this, we model the estimated channel matrices for coordinated and uncoordinated cases, respectively, as $\hat{\mathbf{H}}_{k,l} = \mathbf{H}_{k,l} + \triangle_{k,l}$ and $\hat{\boldsymbol{\Psi}}_{k,m} = \tilde{\mathbf{H}}_{k,m} + \tilde{\triangle}_{k,m}$, where $\triangle_{k,l}$, $\tilde{\triangle}_{k,m}$ are the additive estimation error imposed on the CSI. For the coordinated case, the channel estimation errors are uncorrelated and orthogonal with $\mathbf{H}_{k,l}$, so that their entries are assumed to be i.i.d. zero mean circularly symmetric Gaussian random variables with variance $\sigma_{\triangle}^2$. This also holds for the uncoordinated case (with variance $\sigma_{\tilde{\triangle}}^2$). Therefore, the variances of the channel entries and the estimated channel entries σ_h^2,$\sigma_{\hat{h}}^2$ for coordinated interference, $\sigma_{\tilde{h}}^2$, $\sigma_{\hat{\Psi}}^2$ for uncoordinated interference can be written as

$$\sigma_{\triangle}^2 = (1-\rho^2)\sigma_h^2, \sigma_{\tilde{\triangle}}^2 = (1-\tau^2)\sigma_{\tilde{h}}^2 \quad (2a)$$

$$\sigma_{\hat{h}}^2 = \rho^2\sigma_h^2, \sigma_{\hat{\Psi}}^2 = \tau^2\sigma_{\tilde{h}}^2 \quad (2b)$$

where ρ and τ denote the estimation quality for both cases (e.g., denoting full CSI when $\rho = 1$). Therefore, the received signal at the k-th receiver with imperfect CSI is

$$\boldsymbol{y}_{k-sum} = \boldsymbol{y}_{k-es} - \boldsymbol{W}_k^H \sum_{l=1}^{K} \triangle_{k,l} \boldsymbol{F}_l \boldsymbol{s}_l - \boldsymbol{W}_k^H \sum_{m=1}^{M} \tilde{\triangle}_{k,m} \ \widetilde{\boldsymbol{F}}_m \ \boldsymbol{x}_m \quad (3)$$

where $\boldsymbol{y}_{k-es}$ denotes the received signal at the k-th receiver by using the estimated channel matrices ($\hat{\mathbf{H}}_{k,l}$, $\hat{\boldsymbol{\Psi}}_{k,m}$) in (1).

The main IA goal is that the desired signal spans over all d dimensions, at the same time that the interference subspaces have zero dimension. Therefore, perfect IA holds when the following conditions are satisfied:

$$\boldsymbol{W}_k^H \boldsymbol{H}_{k,l} \boldsymbol{F}_l = 0, \quad (4)$$

$$\text{rank}\left(\boldsymbol{W}_k^H \boldsymbol{H}_{k,k} \boldsymbol{F}_k\right) = d, \quad (5)$$

where we assume that all users intend to achieve the same DoF given by d.

The system $(N_t \times N_r, d)^K$ is proper when the number of variables is equal or smaller than the number of equations, implying in $N_r + N_t - d(K+1) \geq 0$, which has been shown to be necessary but not sufficient in a general case [22]. Nevertheless, when $N_t = N_r$ the necessary condition also becomes sufficient [3], which was latter shown in [23] that the necessary condition can be sufficient even when $N_t \neq N_r$. However, due to the considered sources of uncoordinated interference, even if (4) and (5) are satisfied, the DoF d may not be achieved due to the presence of the uncoordinated interference [23]. Indeed, since the system is improper, the dimension of interference is not zero. Thus, designing precoding and linear receive matrices to minimize the rank of the interference is the main purpose here.

Defining the signal and interference matrices as $\boldsymbol{S}_k$ and $\boldsymbol{J}_k$, $\forall k \in [1, K]$, respectively as

$$\boldsymbol{S}_k \triangleq \boldsymbol{W}_k^H \boldsymbol{H}_{k,k} \boldsymbol{F}_k, \quad (6)$$

$$\boldsymbol{J}_k \triangleq \boldsymbol{W}_k^H \left[\{\boldsymbol{H}_{k,l} \boldsymbol{F}_l\}_{l=1,l\neq k}^{K} \cdots \left\{ \widetilde{\boldsymbol{H}}_{k,m} \widetilde{\boldsymbol{F}}_m \right\}_{m=1}^{M} \right], \quad (7)$$

where $\boldsymbol{J}_k \in \mathbb{R}^{d \times [(K-1)d + A_m]}$ and $A_m = \sum_{m=1}^{M} d_m$. The columns of $\mathbf{S}_k$ span the subspace in which the k-th receiver observes the transmitted signal, and the interference subspace is spanned by the columns of $\mathbf{J}_k$. Besides, DOF of the k-th user can be expressed as $d_k = \text{rank}(\mathbf{S}_k) - \text{rank}(\mathbf{J}_k)$ [4]. We can state (4) and (5) as $\text{rank}(\mathbf{S}_k) = d$ and $\text{rank}(\mathbf{J}_k) = 0$.

Therefore, to maximize DOF of each user (via minimizing the rank of the interference matrix and fixing the rank of the signal matrix), the precoding and receive filter matrices are designed as follows:

$$\min_{\substack{\boldsymbol{W}_k, \boldsymbol{F}_k,\\ k\in[1,K]}} \sum_{k=1}^{K} \text{rank}(\boldsymbol{J}_k), \quad (8a)$$

$$\text{s.t.:} \qquad \text{rank}(\boldsymbol{S}_k) = d. \quad (8b)$$

The minimization problem in (8) involves the rank function of the interference, which is intractable in closed-form. In the literature, the nuclear norm (or l_1 norm of singular values) has been used as a surrogate of the rank function in [4]. The nuclear norm is defined as the sum of the singular values of the matrix, denoted by $\|.\|_* = \sum_{i=1}^{\text{rank}(.)} \delta_i(.)$. Then, since the rank of a semidefinite matrix equals to the number of its non-zero singular values, it is expected that lower rank solutions are obtained by minimizing the nuclear norm. However, although it can be solved efficiently [24], the nuclear norm emphasizes small singular values, while less weight is put on large singular values [25]. When perfect IA cannot be attainable, then the rank of the interference matrix is not zero, a difference between the rank function and nuclear norm appears. Therefore, the number of DoF decreases.

This motivates us to search for other approaches in order to increase the DoF. In the following, we propose two semidefinite programming (SDP)-based rank minimization methods with the ability to enhance the performance of IA in the presence of uncoordinated interference sources.

B. SDP Constrained l_2 Norm Minimization for Perfect CSI

Choosing appropriate weights for the singular values is an important issue to minimize the rank function, especially in presence of uncoordinated interference ($N_r + N_t - d(K+1) \geq 0$ does not hold, and the rank of interference matrix is not zero), and this leads to increase the singular values of $\boldsymbol{J}_k$ (then DoF decreases).

This can be shown via the interpretation of the singular value decomposition (SVD). From linear algebra, the length of the projection of j_i (the i-th row of $\mathbf{J}^H$) onto $\mathbf{v}$ is $||j_i\mathbf{v}||$; thus, the sum of length squared of the projections is $||\mathbf{J}^H\mathbf{v}||^2$. Moreover, the first singular value of $\mathbf{J}^H$ which is denoted by $\mathbf{v}_1$ is a column vector that maximizes the sum of length of the projections onto $\mathbf{v}$ i.e. $\mathbf{v}_1 = \arg\max_{||\mathbf{v}||=1} ||\mathbf{J}^H\mathbf{v}||^2$. Also, $\sigma_1^2 = ||\mathbf{J}^H\mathbf{v}_1||^2$($\sigma_1$ is the first singular value of $\mathbf{J}^H$) is the sum of the squares of the projections of the points to the line determined by $\mathbf{v}_1$ [26]. We can continue this procedure to find all singular vectors (they should be perpendicular to each other). Suppose that we add a new column to $\mathbf{J}$ (for instance one uncoordinated interference source exists in the network), because the cost function that yields $\mathbf{v}_1$ is separable, the singular values of $\mathbf{J}^H$ increases. Therefore, existence of uncoordinated source leads to increase the singular values of $\mathbf{J}$.

To minimize the rank of interference matrix, minimizing l_2 norm of its singular values is a good candidate for such scenarios since it sets higher weights for larger singular values. Indeed, in l_p norm approximation, a penalty function is defined as $\phi_p(u) = |u|^p$, where u is denoted as residual. Then, comparing the l_1 and l_2 norms, for small u we will have $\phi_1(u) \gg \phi_2(u)$, meaning that the l_1 norm approximation [4] will put relatively larger emphasis on small residuals compared to the l_2 norm approximation. However, for large u we will have $\phi_2(u) \gg \phi_1(u)$; thus, the l_1 norm puts less weights on large residuals in comparison with l_2 norm approximation [25]. Consequently, l_2 norm yields fewer large singular values than the nuclear norm [4]. This characteristic is beneficial to achieve higher DoF with uncoordinated interference. Thus, using the l_2 norm to approximate the rank function we have

$$\min_{\substack{\boldsymbol{W}_k, \boldsymbol{F}_k,\\ k\in[1,K]}} \sum_{k=1}^{K} \mathrm{tr}(\boldsymbol{Z}_k), \tag{9a}$$
$$\text{s.t.:} \quad \mathrm{rank}(\boldsymbol{S}_k) = d, \tag{9b}$$
$$\boldsymbol{Z}_k = \boldsymbol{J}_k^H \boldsymbol{J}_k, \tag{9c}$$

recalling that $\boldsymbol{S}_k$ and $\boldsymbol{J}_k$ are respectively given by (6) and (7). Moreover, since (9c) is not a constraint of a convex optimization problem, we use the Schur complement to convert it into a linear matrix inequality (LMI). To this end, we relax (9c) to $\mathbf{Z}_k \succeq \boldsymbol{J}_k^H \boldsymbol{J}_k$. Using the schur complement, this constraint can be expressed as follows $\begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{J}_k \\ \boldsymbol{J}_k^H & \boldsymbol{Z}_k \end{pmatrix} \succeq \mathbf{0}$ [27]. so that

$$\min_{\substack{\boldsymbol{W}_k, \boldsymbol{F}_k,\\ k\in[1,K]}} \sum_{k=1}^{K} \mathrm{tr}(\boldsymbol{Y}_k), \tag{10a}$$
$$\text{s.t.:} \quad \mathrm{rank}(\boldsymbol{S}_k) = d, \tag{10b}$$
$$\boldsymbol{Y}_k \succeq \mathbf{0}_{\zeta\times\zeta}, \tag{10c}$$
$$\boldsymbol{Y}_k = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{J}_k \\ \boldsymbol{J}_k^H & \boldsymbol{Z}_k \end{pmatrix}, \tag{10d}$$

with $\zeta = Kd + A_m$.

Yet, the rank constraint in (10b) does not imply a convex feasible set. To address this issue, we restrict the signal matrix to positive definite matrices, which are full rank (they satisfy (10b)). In order to satisfy this criterion, we resort to the following lemma.

Lemma 1. *$\boldsymbol{S}_k$ is positive definite and full rank when*

$$\boldsymbol{S}_k - \gamma\,\boldsymbol{I}_d \succeq \mathbf{0}_{d\times d}, \tag{11}$$

where $0 < \gamma \ll 1$.

Proof. If the scalar $\boldsymbol{b}^T\boldsymbol{A}\boldsymbol{b}$ is positive for any given non-zero column vector $\boldsymbol{b}$, then the symmetric matrix $\boldsymbol{A}$ is positive definite. Due to the fact that (11) is positive semidefinite, we have

$$\boldsymbol{b}^T(\boldsymbol{S}_k - \gamma\boldsymbol{I}_d)\boldsymbol{b} \geq 0, \tag{12}$$
$$\boldsymbol{b}^T\boldsymbol{S}_k\boldsymbol{b} \geq \gamma\boldsymbol{b}^T\boldsymbol{I}_d\boldsymbol{b} > 0. \tag{13}$$

Then, since $\boldsymbol{I}_d$ is positive definite and $\gamma > 0$, $\boldsymbol{S}_k$ is also positive definite and full rank. □

Consequently, by using Lemma 1 we can state our optimization problem as follows:

$$\min_{\substack{\boldsymbol{W}_k, \boldsymbol{F}_k,\\ k\in[1,K]}} \sum_{k=1}^{K} \mathrm{tr}(\boldsymbol{Y}_k), \tag{14a}$$
$$\text{s.t.:} \boldsymbol{S}_k - \gamma\,\boldsymbol{I}_d \succeq \mathbf{0}_{d\times d} \tag{14b}$$
$$\boldsymbol{Y}_k \succeq \mathbf{0}_{\zeta\times\zeta}, \tag{14c}$$
$$\boldsymbol{Y}_k = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{J}_k \\ \boldsymbol{J}_k^H & \boldsymbol{Z}_k \end{pmatrix}, \tag{14d}$$

Lemma 2. *The dual form of (14) at each iteration of the alternating minimization approach can be expressed as follows:*

$$\max \sum_{k=1}^{K} \mathrm{tr}\left(\gamma\boldsymbol{C}_{1,k} - \boldsymbol{D}_{3,k}\right) \tag{15a}$$
$$\textit{s.t.:} \quad \boldsymbol{I}_\zeta - \boldsymbol{C}_{2,k} + \boldsymbol{D}\boldsymbol{D}_{1,k} + \boldsymbol{D}\boldsymbol{D}_{2,k} = \mathbf{0}_{\zeta\times\zeta}, \tag{15b}$$
$$\boldsymbol{D}_{1,k} - \boldsymbol{C}_{1,k} = \mathbf{0}_{d\times d}, \tag{15c}$$
$$\boldsymbol{C}_{1,k} \succeq \mathbf{0}_{d\times d}, \tag{15d}$$
$$\boldsymbol{C}_{2,k} \succeq \mathbf{0}_{\zeta\times\zeta}, \tag{15e}$$
$$\boldsymbol{H}_{k,k}\boldsymbol{F}_k\boldsymbol{D}_{1,k} + \boldsymbol{\Theta}\boldsymbol{D}_{2,k} = \mathbf{0}_{N_r\times d}, \tag{15f}$$

where $\boldsymbol{C}_{1,k}$ and $\boldsymbol{C}_{2,k}$ are Lagrange multipliers associated with the inequality constraints, $\boldsymbol{D}_{1,k}$, $\boldsymbol{D}_{2,k}$ and $\boldsymbol{D}_{3,k}$ are Lagrange multipliers associated with the equality constraints,

while $\boldsymbol{\Theta} = \left[\{\boldsymbol{H}_{k,l}\boldsymbol{F}_l\}_{l=1,l\neq k}^{K} \cdots \left\{\widetilde{\boldsymbol{H}}_{k,m}\widetilde{\boldsymbol{F}}_m\right\}_{m=1}^{M}\right]$, $\boldsymbol{DD}_{1,k} = \begin{pmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times(\zeta-d)} \\ \boldsymbol{D}_{2,k} & \mathbf{0}_{(\zeta-d)\times(\zeta-d)} \end{pmatrix}$ *and*
$\boldsymbol{DD}_{2,k} = \begin{pmatrix} \boldsymbol{D}_{3,k} & \mathbf{0}_{d\times(\zeta-d)} \\ \mathbf{0}_{(\zeta-d)\times d} & \mathbf{0}_{(\zeta-d)\times(\zeta-d)} \end{pmatrix}$.

In order to obtain the dual problem of (14) at each iteration of the alternating minimization approach, which is a single variable problem, we employ the Lagrange dual function. Please see Appendix A.

Since, we replace (10b) by (11), our optimization problem is convex and SDP at each iteration of the alternating minimization approach. Therefore, strong duality holds (means that the optimal values of the primal and dual problems are equal) at each iteration of the alternating minimization approach.

Let us now introduce a matrix $\boldsymbol{\Upsilon}$ into Lemma 1 in order to consider new possible solutions that have been disregarded in [4], [8]. As we stated before, considering the new solutions can lead to get a better result. Thus, we have:

$$\boldsymbol{S}_k + \boldsymbol{\Upsilon} - \gamma\, \boldsymbol{I}_d \succeq \mathbf{0}_{d\times d}. \tag{16}$$

If $\boldsymbol{\Upsilon}$ satisfies $\mathrm{tr}(\boldsymbol{C}_{1,k}\boldsymbol{\Upsilon}) \geq 0$, from Proposition 1, we conclude that the optimal value of the dual problem associated with the constraint in (16) is smaller than the case when $\boldsymbol{\Upsilon} = \mathbf{0}$. Therefore, adding the matrix $\boldsymbol{\Upsilon}$ can decrease the rank of the interference matrices compared to the optimization problem in (14). Consequently, the IA scheme may achieve higher DoF.

Proposition 1. *Introducing the matrix $\boldsymbol{\Upsilon}$ using* (16) *modifies the objective function of the dual problem as follows:*

$$\max\ \mathrm{tr}\Big(\boldsymbol{C}_{1,k}(\gamma \boldsymbol{I}_d - \boldsymbol{\Upsilon}) - \boldsymbol{D}_{3,k}\Big). \tag{17}$$

Proof. Please see Appendix B. □

Proposition 2. *Let the optimal values of the problem in* (14)*, associated with the constraints in* (11) *and* (16) *be P^* and $P^*(\boldsymbol{\Upsilon})$, respectively. Then, for any $\boldsymbol{\Upsilon}$ we have that*

$$P^*(\boldsymbol{\Upsilon}) \geq P^* - \mathrm{tr}(\boldsymbol{C}_{1,k}^*\boldsymbol{\Upsilon}). \tag{18}$$

Proof. Please see Appendix C. □

Thus, in order to be as close as possible to the lower bound of $P^*(\boldsymbol{\Upsilon})$, $\boldsymbol{\Upsilon}$ should be relatively small. This implies that, in order to obtain lower rank solutions for the interference matrix, we choose small values for the entries of $\boldsymbol{\Upsilon}$ in a way that the condition $\mathrm{tr}(\boldsymbol{C}_{1,k}\boldsymbol{\Upsilon}) \geq 0$ is satisfied. Since $\boldsymbol{C}_{1,k}$ is positive semidefinite, $\mathbf{1}_d$ satisfies $\mathrm{tr}(\boldsymbol{C}_{1,k}\boldsymbol{\Upsilon}) \geq 0$. Indeed, $\mathrm{tr}(\boldsymbol{C}_{1,k}\boldsymbol{\Upsilon})$ equals the summation of the diagonal elements of $\boldsymbol{C}_{1,k}\boldsymbol{\Upsilon}$. When $\boldsymbol{\Upsilon} = \mathbf{1}_d$, $\mathrm{tr}(\boldsymbol{C}_{1,k}\boldsymbol{\Upsilon})$ equals the summation of all elements of $\boldsymbol{C}_{1,k}$ which is equivalent to $\boldsymbol{a}_1^T\boldsymbol{C}_{1,k}\boldsymbol{a}_1$ where $\boldsymbol{a}_1$ is a vector that all elements of it equals one. Besides, $\boldsymbol{C}_{1,k}$ is positive semidefinite which leads to $\boldsymbol{a}_1^T\boldsymbol{C}_{1,k}\boldsymbol{a}_1 \geq 0$. Thus, we can conclude that $\mathrm{tr}(\boldsymbol{C}_{1,k}\boldsymbol{\Upsilon}) \geq 0$. Therefore, the matrix $\boldsymbol{\Upsilon}$ is chosen as follows:

$$\boldsymbol{\Upsilon} = \upsilon\, \mathbf{1}_d, \tag{19}$$

where υ is a positive constant. Although (16) does not guarantee that $\boldsymbol{S}_k$ is positive definite, our results show that by choosing sufficiently small values for υ provides full rank $\boldsymbol{S}_k$. Either way, if $\boldsymbol{S}_k$ is not full rank, we decrease υ step by step until a full rank matrix is obtained. Therefore, one can conclude that using $\boldsymbol{\Upsilon}$ expands the feasible set of the optimization problem, which enables us to obtain lower optimal values.

Algorithm 1 Alternating minimization approach to solve the optimization problem in (10)

1: Choose an arbitrary matrix $\boldsymbol{W}_k$, $k \in [1, K]$, ϵ and $n = 1$
2: fix $\boldsymbol{W}_k$ and solve (20) with respect to $\boldsymbol{F}_k$
3: fix $\boldsymbol{F}_k$ and solve (20) with respect to $\boldsymbol{W}_k$
4: **if** $\mathrm{rank}(\boldsymbol{S}_k) \neq d$ **then**
5: $\quad \upsilon \to \upsilon/2$
6: **else**
7: $\quad$ end algorithm
8: **end if**
9: $n = n + 1$
10: Until $f(n) - f(n-1) \leq \epsilon$

Finally, our proposed optimization problem can be state as:

$$\min_{\substack{\boldsymbol{W}_k, \boldsymbol{F}_k,\\ k\in[1,K]}} \sum_{k=1}^{K} \mathrm{tr}(\boldsymbol{Y}_k), \tag{20a}$$

$$\text{s.t.:} \boldsymbol{S}_k + \boldsymbol{\Upsilon} - \gamma\, \boldsymbol{I}_d \succeq \mathbf{0}_{d\times d} \tag{20b}$$

$$\boldsymbol{Y}_k \succeq \mathbf{0}_{\zeta\times\zeta}, \tag{20c}$$

$$\boldsymbol{Y}_k = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{J}_k \\ \boldsymbol{J}_k^H & \boldsymbol{Z}_k \end{pmatrix}, \tag{20d}$$

since (20) has two variables, we use an alternating minimization approach [28] to solve it which is detailed in Algorithm 1 where $f(.)$ is the cost function of the optimization problem and ϵ is a convergence threshold.

C. SDP Constrained l_2 Norm Minimization for ImPerfect CSI

As we can observe from (3), in the imperfect CSI case, we have additional terms at the receivers, due to the estimation noise, which scales up with the power of transmitters (especially due to the term $\boldsymbol{W}_k^H \sum_{l=1}^{K} \triangle_{k,l}\, \boldsymbol{F}_l$). This can reduce DoF of the network, making it interference-limited. Thus, to tackle this issue, we can minimize the power of the noise caused by imperfect CSI at each receiver expressed as $\sum_{l=1}^{K} \sigma_{\triangle}^2 \mathrm{tr}(\mathbf{W}_k^H \mathrm{tr}(\mathbf{F}_l\mathbf{F}_l^H)\mathbf{W}_k)$ for the k-th receiver. The power of enhanced noise does not imply a cost function of a convex optimization problem. One approach to minimize it could be to fix $\mathbf{W}_k$ and minimize $\mathrm{tr}(\mathbf{F}_l\mathbf{F}_l^H)$ in a first moment, then to fix $\mathbf{F}_l$ and minimize $\mathrm{tr}(\mathbf{W}_k\mathbf{W}_k^H)$ secondly. Since the former terms are non-convex, similar to (9) we can use the Schur complement to relax them. Consequently, we have the following optimization problems for the imperfect CSI case:

$$\min_{\substack{\boldsymbol{F}_k \\ k\in[1,K]}} \sum_{k=1}^{K} \operatorname{tr}(\boldsymbol{Y}_k) + \sum_{k=1}^{K} \operatorname{tr}(\boldsymbol{R}_k), \tag{21a}$$

$$\text{s.t.:} \boldsymbol{S}_k + \boldsymbol{\Upsilon} - \gamma \boldsymbol{I}_d \succeq \boldsymbol{0}_{d\times d} \tag{21b}$$

$$\boldsymbol{Y}_k \succeq \boldsymbol{0}, \boldsymbol{R}_k \succeq \boldsymbol{0} \tag{21c}$$

$$\boldsymbol{Y}_k = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{J}_k \\ \boldsymbol{J}_k^H & \boldsymbol{Z}_k \end{pmatrix}, \boldsymbol{R}_k = \begin{pmatrix} \boldsymbol{I} & \mathbf{F}_k \\ \mathbf{F}_k^H & \boldsymbol{R}'_k \end{pmatrix} \tag{21d}$$

and

$$\min_{\substack{\boldsymbol{W}_k \\ k\in[1,K]}} \sum_{k=1}^{K} \operatorname{tr}(\boldsymbol{Y}_k) + \sum_{k=1}^{K} \operatorname{tr}(\boldsymbol{\Gamma}_k), \tag{22a}$$

$$\text{s.t.:} \boldsymbol{S}_k + \boldsymbol{\Upsilon} - \gamma \boldsymbol{I}_d \succeq \boldsymbol{0}_{d\times d} \tag{22b}$$

$$\boldsymbol{Y}_k \succeq \boldsymbol{0}, \boldsymbol{\Gamma}_k \succeq \boldsymbol{0} \tag{22c}$$

$$\boldsymbol{Y}_k = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{J}_k \\ \boldsymbol{J}_k^H & \boldsymbol{Z}_k \end{pmatrix}, \boldsymbol{\Gamma}_k = \begin{pmatrix} \boldsymbol{I} & \mathbf{W}_k \\ \mathbf{W}_k^H & \boldsymbol{\Gamma}'_k \end{pmatrix} \tag{22d}$$

where $\mathbf{R}'_k \succeq \mathbf{F}_k^H \mathbf{F}_k$ and $\boldsymbol{\Gamma}'_k \succeq \mathbf{W}_k^H \mathbf{W}_k$. The procedure to solve (21)-(22) is similar to that of (20).

D. Schatten-p Norm Minimization

Alternatively, the Schatten-p norm may also be used as a surrogate of the rank operator. The Schatten-p norm, with $p \in (0,1]$, is defined as

$$f_p(\boldsymbol{A}) = \sum_{i>1} (\delta_i(\boldsymbol{A}))^p, \tag{23}$$

which equals to the nuclear norm when $p = 1$. Hence, the nuclear norm is a special case of the Schatten-p norm. In addition to, when p tends to zero, $f_p(\boldsymbol{A}) \to \operatorname{rank}(\boldsymbol{A})$. This property helps us to approximate the sum of the number of non-zero singular values, instead of their magnitudes as in [4], and in comparison with [8], it enables us to treat the singular values closer to the rank function. Also, it suggests that the Schatten-p norm may find low-rank solutions when p is small.

However, to avoid the non-differentiability of (23), we resort to [29] so that the objective function to be minimized can be written as

$$\min_{\boldsymbol{W}_k, \boldsymbol{F}_k} \sum_{k=1}^{K} \sum_{i=1}^{d} (\delta_i(\boldsymbol{J}_k) + \xi)^p, \tag{24}$$

where ξ is a constant with a small positive value. Yet, (24) is not convex, so that we employ a Taylor expansion to linearize it. Then, at the l-th iteration the expansion of (24) is given by

$$\left(\delta_i(\boldsymbol{J}_k^l) + \xi\right)^p + \left(\frac{p}{\left(\delta_i(\boldsymbol{J}_k^l) + \xi\right)^{1-p}}\right) \left(\delta_i(\boldsymbol{J}_k) - \delta_i(\boldsymbol{J}_k^l)\right). \tag{25}$$

Here, $\boldsymbol{J}_k^l$ is fixed and can be removed. As a result, the objective function simplifies to

$$\min_{\boldsymbol{W}_k, \boldsymbol{F}_k} \sum_{k=1}^{K} \sum_{i=1}^{d} \frac{p\, \delta_i(\boldsymbol{J}_k)}{\left(\delta_i(\boldsymbol{J}_k^l) + \xi\right)^{1-p}}. \tag{26}$$

Algorithm 2 Iterative algorithm to solve (27)

Choose an arbitrary matrix $\boldsymbol{W}_k$, $k \in [1, K]$, ϵ and $l = 1$
$\boldsymbol{G}_k^0 \leftarrow \boldsymbol{I}_d$
for $n = 1 : n_{\max}$ **do**
  fix $\boldsymbol{W}_k$ and solve (27) with respect to $\boldsymbol{F}_k$
  fix $\boldsymbol{F}_k$ and solve (27) with respect to $\boldsymbol{W}_k$
  if $\operatorname{rank}(\boldsymbol{S}_k) \neq d$ **then**
    $\upsilon \to \upsilon/2$
  else
    end algorithm
  end if
end for
$l = l + 1$
update $\boldsymbol{G}_k^{l+1}$ according to $\boldsymbol{J}_k^l$
Until $f(l) - f(l-1) \leq \epsilon$

Proposition 3. *The complete optimization problem using the Schatten-p norm is given by:*

$$\min_{\boldsymbol{W}_k, \boldsymbol{F}_k} \sum_{k=1}^{K} \left\| \boldsymbol{G}_k^l \boldsymbol{J}_k \right\|_*, \tag{27a}$$

$$s.t.: \quad \boldsymbol{S}_k + \boldsymbol{\Upsilon} - \gamma \boldsymbol{I}_d \succeq \boldsymbol{0}_{d\times d}, \tag{27b}$$

where $\boldsymbol{G}_k^l = \boldsymbol{\Lambda}_k \boldsymbol{\Phi}_k^l \boldsymbol{\Lambda}_k^H$ is the weight matrix and $\boldsymbol{\Lambda}_k$ is obtained by the singular value decomposition of $\boldsymbol{J}_k$. Moreover, $\boldsymbol{\Phi}_k^l \in \mathbb{R}^{d\times d}$ is a diagonal matrix whose i-th diagonal element is given by $\frac{p}{\left(\delta_i(\boldsymbol{J}_k^l) + \xi\right)^{1-p}}$.

Proof. Please see Appendix D. □

An iterative approach to solve (27) is formalized by Algorithm 2 where $f(.)$ is the cost function of the optimization problem and ϵ is a convergence threshold. For the imperfect case, we can use (21)-(22) (we should just use (27a) instead of $\sum_{k=1}^{K} \operatorname{tr}(\boldsymbol{Y}_k)$).

III. The Proposed Method for Relay-Aided MIMO Interference Channel

Now, we extend the previous system model to the relay-aided MIMO interference channels. Let us assume that X half-duplex AF relays assist the communication among the K user pairs. This system can be represented by $(N_t \times N_r, d)^K + (N_x)^X$, where N_x is the number of antennas at each relay. Moreover, we assume no direct link between transmitters and receivers, so that we denote the channel between the k-th transmitter and the x-th relay by $\boldsymbol{H}_{x,k} \in \mathbb{R}^{N_x \times N_t}$, while the channel between the x-th relay and the k-th receiver is $\boldsymbol{G}_{k,x} \in \mathbb{R}^{N_r \times N_x}$.

Due to the half-duplex assumption, the communication process contains two phases. First, each x-th relay receives $\boldsymbol{y}_{x,r} = \sum_{k=1}^{K} \boldsymbol{H}_{x,k} \boldsymbol{F}_k \boldsymbol{s}_k + \boldsymbol{n}_x$, where $\boldsymbol{F}_k$ is the precoding matrix used by the k-th transmitter, and $\boldsymbol{n}_x$ is the noise vector with zero mean and covariance $\sigma^2 \boldsymbol{I}_{N_x}$. Then, in the second phase, the relays amplify their received signals and forward them to the receivers. After applying the linear receive filter $\boldsymbol{W}_k \in \mathbb{R}^{N_r \times d}$, each k-th receiver observes

$$\boldsymbol{y}_k = \sum_{x=1}^{X} \boldsymbol{W}_k^H \boldsymbol{G}_{k,x} \boldsymbol{U}_x \boldsymbol{R}_{x,k} + \boldsymbol{W}_k^H \sum_{l=1, l\neq k}^{K} \sum_{x=1}^{X} \boldsymbol{G}_{k,x} \boldsymbol{U}_x \boldsymbol{R}_{x,l} + \boldsymbol{W}_k^H \sum_{x=1}^{X} \boldsymbol{G}_{k,x} \boldsymbol{U}_x \boldsymbol{n}_x + \boldsymbol{W}_k^H \boldsymbol{f}_k, \quad (28)$$

where $\boldsymbol{U}_x \in \mathbb{R}^{N_x \times N_x}$, $\forall x \in [1, X]$, is the processing matrix at the x-th relay, $\boldsymbol{R}_{x,l} = \boldsymbol{H}_{x,l}\, \boldsymbol{F}_l\, \boldsymbol{s}_l$, and $\boldsymbol{f}_k$ is the zero-mean white Gaussian noise with covariance $\sigma^2 \boldsymbol{I}_{N_r}$.

Moreover, observing (28) we remark that its second term represents the interference produced by the other transmitters at each receiver. In addition, notice also that $\boldsymbol{n}_x$ in the third term of (28) is amplified by the relays along with the desired signal, which we denote by *enhanced noise*. This issue cannot be overlooked once it provides a serious obstacle in achieving higher DoF [14]. Therefore, our approach for this AF relay-aided MIMO interference scenario is to minimize the effect of both the interference and the enhanced noise in a unified design[1].

First, let us define the transmit power of the x-th relay and of the k-th transmitter as

$$PR_x = \sum_{k=1}^{K} \mathrm{tr}\left(\boldsymbol{U}_x\, \boldsymbol{H}_{x,k}\, \boldsymbol{F}_k\, \boldsymbol{F}_k^H\, \boldsymbol{H}_{x,k}^H\, \boldsymbol{U}_x^H\right) + \sigma^2 \boldsymbol{I}_{N_x} \mathrm{tr}\left(\boldsymbol{U}_x\, \boldsymbol{U}_x^H\right), \quad (29)$$

$$PU_k = \mathrm{tr}\left(\boldsymbol{F}_k\, \boldsymbol{F}_k^H\right), \quad (30)$$

respectively, while the signal and interference matrices $(\boldsymbol{S}_k, \boldsymbol{J}_k)$ are given by

$$\boldsymbol{S}_k \triangleq \sum_{x=1}^{X} \boldsymbol{W}_k^H\, \boldsymbol{G}_{k,x}\, \boldsymbol{U}_x\, \boldsymbol{H}_{x,k}\, \boldsymbol{F}_k, \quad (31)$$

$$\boldsymbol{J}_k \triangleq \boldsymbol{W}_k^H \left[\{\boldsymbol{G}_{k,x}\, \boldsymbol{U}_x\, \boldsymbol{H}_{x,l}\, \boldsymbol{F}_l\}_{l=1, l\neq k, x=1}^{l=K, x=X}\right], \quad (32)$$

with $\boldsymbol{J}_k \in \boldsymbol{C}^{d \times [(K-1)dX]}$.

Furthermore, the sum power of the enhanced noise can be written as

$$N = \sum_{k=1}^{K} \sum_{x=1}^{X} \sigma^2 \mathrm{tr}\left(\boldsymbol{W}_k^H\, \boldsymbol{G}_{k,x}\, \boldsymbol{U}_x\, \boldsymbol{U}_x^H\, \boldsymbol{G}_{k,x}^H\, \boldsymbol{W}_k\right), \quad (33)$$

The optimization problem to minimize the effects of interference (the rank of interference) and the enhanced noise (the sum-power of enhanced noise) by designing $\mathbf{W}_k, \mathbf{F}_k$ and $\mathbf{U}_x$ can be staded as follows :

$$\begin{aligned} \min_{\substack{\boldsymbol{W}_k, \boldsymbol{F}_k, \boldsymbol{U}_x, \\ k\in[1,K],\, x\in[1,X]}} \quad & \sum_{k=1}^{K} \mathrm{rank}\,(\boldsymbol{J}_k) + N, && (34a)\\ \text{s.t.:} \quad & PR_x = P_1, && (34b)\\ & PU_k = P_2, && (34c)\\ & \mathrm{rank}\,(\boldsymbol{S}_k) = d, && (34d) \end{aligned}$$

where P_1 and P_2 are the amount of power available at each relay and transmitter, respectively. In (34), we aim to minimize the rank of the interference matrix and the sum power of the enhanced noise.

In this section, we propose a weighted-sum method to solve (34). We start with the design of the precoding matrix $\boldsymbol{F}_k$, for which we employ the Schatten-p norm as a surrogate of the rank function in (34a), which implies in (34d) being replaced by (16). Moreover, since $\boldsymbol{F}_k$ does not affect the sum-power of the enhanced noise, the term N in (34a) can be removed here. In addition, (34b) and (34c) are not constraints of a convex optimization problem, for which we use the Schur complement to turn them into the convex constraints. Therefore,

$$\begin{aligned} \min_{\boldsymbol{F}_k, k\in[1,K]} \quad & \sum_{k=1}^{K} \left\| \boldsymbol{G}_k^l \boldsymbol{J}_k \right\|_*, && (35a)\\ \text{s.t.:} \quad & \boldsymbol{FF}_k = \begin{pmatrix} \boldsymbol{I}_{N_t} & \boldsymbol{F}_k^H \\ \boldsymbol{F}_k & \boldsymbol{F}_{kk} \end{pmatrix}, && (35b)\\ & \boldsymbol{QQ}_{x,k} = \begin{pmatrix} \boldsymbol{I}_d & \boldsymbol{QW}_{x,k}^H \\ \boldsymbol{QW}_{x,k} & \boldsymbol{QA}_{x,k} \end{pmatrix}, && (35c)\\ & \boldsymbol{FF}_k,\, \boldsymbol{QQ}_{x,k} \succeq \boldsymbol{0}, && (35d)\\ & \mathrm{tr}(\boldsymbol{FF}_k) = P_2 + N_t, && (35e)\\ & \sum_{k=1}^{K} \mathrm{tr}(\boldsymbol{QQ}_{x,k}) = P_1 + d - \mathrm{tr}(\boldsymbol{U}_x \boldsymbol{U}_x^H), && (35f)\\ & \boldsymbol{S}_k - \gamma\, \boldsymbol{I}_d \succeq \boldsymbol{0}_{d\times d}, && (35g) \end{aligned}$$

where $\boldsymbol{F}_{kk} \succeq \boldsymbol{F}_k\, \boldsymbol{F}_k^H$, $\boldsymbol{QW}_{x,k} = \boldsymbol{U}_x\, \boldsymbol{H}_{x,k}\, \boldsymbol{F}_k$, and $\boldsymbol{QA}_{x,k} \succeq \boldsymbol{QW}_{x,k}^H \boldsymbol{QW}_{x,k}$. Then, after solving (35) we still need to normalize $\boldsymbol{F}_k$ in order to satisfy the power constraint in (34c), so that we define $\widehat{\boldsymbol{F}_k} = \alpha \boldsymbol{F}_k$ as the new precoding matrix[2], where $\alpha = \sqrt{\frac{\mathrm{tr}(\boldsymbol{F}_{kk})}{\mathrm{tr}(\boldsymbol{F}_k \boldsymbol{F}_k^H)}}$.

Next, we determine the processing matrices at the relays ($\boldsymbol{U}_x$), for which we need to take the sum-power of the enhanced noise (N) into account. Then, since N and the constraint in (34b) are not defined according to a convex optimization problem, we again resort to the Schur complement

[1]Notice that the enhanced noise in the relay-aided scenario can also be seen as a source of single-hop interference from uncoordinated relays which impacts receivers, as that considered in Section II-A.

[2]Notice that the Schur complement relaxes (34b) and (34c) in order to solve the optimization problem. Therefore, in order to satisfy these constraints we still need to normalize the obtained solution.

in order to rewrite the optimization problem as

$$\min_{\boldsymbol{U}_x, x\in[1,X]} \sum_{k=1}^{K} \left\| \boldsymbol{G}_k^l \boldsymbol{J}_k \right\|_* + \sum_{k=1}^{K}\sum_{x=1}^{X} \sigma^2 \operatorname{tr}(\boldsymbol{T}_{x,k}), \tag{36a}$$

$$\text{s.t.:} \quad \boldsymbol{T}_{x,k} = \begin{pmatrix} \boldsymbol{I}_{N_t} & \boldsymbol{TW}_{x,k}^H \\ \boldsymbol{TW}_{x,k} & \boldsymbol{TA}_{x,k} \end{pmatrix}, \tag{36b}$$

$$\boldsymbol{QQ}_{x,k} = \begin{pmatrix} \boldsymbol{I}_d & \widehat{\boldsymbol{QW}}_{x,k}^H \\ \widehat{\boldsymbol{QW}}_{x,k} & \boldsymbol{QA}_{x,k} \end{pmatrix}, \tag{36c}$$

$$\boldsymbol{C}_x = \begin{pmatrix} \boldsymbol{N}_x & \boldsymbol{U}_x^H \\ \boldsymbol{U}_x & \boldsymbol{U}_{xx} \end{pmatrix}, \tag{36d}$$

$$\sum_{k=1}^{K} \operatorname{tr}(\boldsymbol{QQ}_{x,k}) + \operatorname{tr}(\boldsymbol{C}_x) = P_1 + d + N_x, \tag{36e}$$

$$\boldsymbol{T}_{x,k},\ \boldsymbol{QQ}_{x,k},\ \boldsymbol{C}_x \succeq \boldsymbol{0}, \tag{36f}$$

$$\boldsymbol{S}_k - \gamma \boldsymbol{I}_d \succeq \boldsymbol{0}_{d\times d}, \tag{36g}$$

where $\boldsymbol{TW}_{x,k} = \boldsymbol{W}_k^H \boldsymbol{G}_{k,x} \boldsymbol{U}_x$, $\boldsymbol{TA}_{x,k} \succeq \boldsymbol{TW}_{x,k}^H \boldsymbol{TW}_{x,k}$, $\widehat{\boldsymbol{QW}}_{x,k} = \boldsymbol{U}_x \boldsymbol{H}_{x,k} \widehat{\boldsymbol{F}}_k$, and $\boldsymbol{U}_{xx} \succeq \boldsymbol{U}_x^H \boldsymbol{U}_x$.

Then, using the same rationale to obtain the linear receive matrix $\boldsymbol{W}_k$ we have

$$\min_{\boldsymbol{W}_k, k\in[1,K]} \sum_{k=1}^{K} \left\| \boldsymbol{G}_k^l \boldsymbol{J}_k \right\|_* + \sum_{k=1}^{K}\sum_{x=1}^{X} \sigma^2 \operatorname{tr}(\boldsymbol{T}_{x,k}), \tag{37a}$$

$$\text{s.t.:} \quad \boldsymbol{T}_{x,k} = \begin{pmatrix} \boldsymbol{I}_{N_t} & \boldsymbol{TW}_{x,k}^H \\ \boldsymbol{TW}_{x,k} & \boldsymbol{TA}_{x,k} \end{pmatrix}, \tag{37b}$$

$$\boldsymbol{T}_{x,k} \succeq \boldsymbol{0}, \tag{37c}$$

$$\boldsymbol{S}_k - \gamma \boldsymbol{I}_d \succeq \boldsymbol{0}_{d\times d}. \tag{37d}$$

Since (36) and (37) are bi-objective optimization problems, there are multiple efficient (optimal) points that can optimize them. The weighted-sum is a technique to find efficient points. In this method, a constant weight is allotted to each objective. Nevertheless, the weighted-sum cannot find the whole set of efficient solutions, and the solutions are not uniformly distributed [30]. Therefore, in this paper we propose a modified weighted-sum method inspired in [30] in order to find a larger set of efficient optimal points.

The modified weighted-sum method changes weights adaptively, rather than increases weights with a fixed step as in the traditional approach, which incurs in additional inequality constraints. For instance, Fig. 1 illustrates a bi-objective optimization problem, where f_1 and f_2 represent the two objective functions, and whose Pareto front is a convex region with non-uniform curvature. By using the traditional weighted-sum method we find a few optimal solutions illustrated by the black dots in the flat region of Fig. 1, but not all, regardless of the size of the step of the weighting factor. However, by using the modified weighted-sum method [30] we start with the traditional approach using a large step size for the weighting factor. Then, by calculating the distances between neighbour solutions, denoted by μ, regions for further refinement are identified, which allows us to enlarge the set of efficient solutions, illustrated by the white dots in Fig. 1.

Therefore, we reformulate the objective functions in (36a) and (37a) as

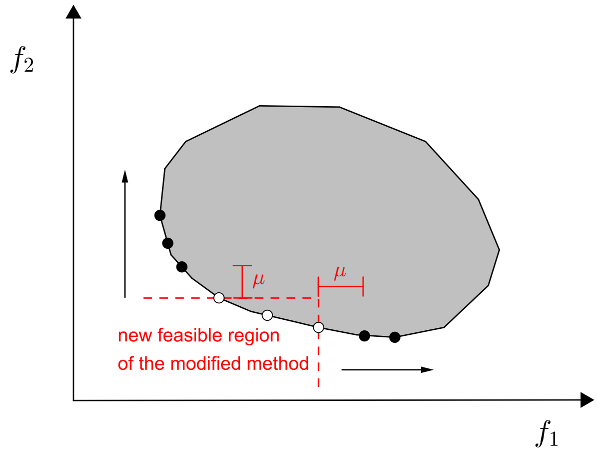


Fig. 1. Modified weighted-sum method in a convex Pareto front. The black dots are the solutions obtained with the traditional approach, while the white dots are available due to the additional refinement of the modified approach.

$$\min\ w_1 \sum_{k=1}^{K} \left\| \boldsymbol{G}_k^l \boldsymbol{J}_k \right\|_* + \sum_{k=1}^{K}\sum_{x=1}^{X} w_2 \sigma^2 \operatorname{tr}(\boldsymbol{T}_{x,k}), \tag{38}$$

where w_1, w_2 are constant weights such that $\sum_{i=1}^{2} w_i = 1$, $w_i > 0$. Moreover, due to the choice of the arbitrary offset distance μ, two extra constraints must be imposed on (36)-(37). Thus:

$$w_1 \left\| \boldsymbol{G}_k^l \boldsymbol{J}_k \right\|_* \leq Z_1^x - \mu \cos(\theta), \tag{39}$$

$$w_2 \operatorname{tr}(\boldsymbol{T}_{x,k}) \leq Z_2^y - \mu \sin(\theta), \tag{40}$$

where $\theta = \tan^{-1}\left(-\frac{Z_1^y - Z_2^y}{Z_1^x - Z_2^x}\right)$, with Z_1^x and Z_1^y representing the respective positions of $\left\| w_1 \boldsymbol{G}_k^l \boldsymbol{J}_k \right\|_*$ and $\operatorname{tr}(w_2 \boldsymbol{T}_{x,k})$ in the Pareto front, calculated when $w_1 = \alpha_1$ and $w_2 = 1 - \alpha_1$, while Z_2^x and Z_2^y are the same positions recalculated when $w_1 = 1 - \alpha_1$ and $w_2 = \alpha_1$.

The proposed modified weighted-sum method is detailed in Algorithm 3, where notice that a set of weights $\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$ is employed in order to expand the set of solutions, once different weights yield different efficient points. Moreover, similarly to previous sections we use an alternating minimization approach to solve (36)-(37). In addition, due to the relaxation caused by the Schur complement, we still need to normalize the relay processing matrix in order to satisfy (34b). Therefore, we define $\widehat{\boldsymbol{U}}_x = \beta \boldsymbol{U}_x$ as the new relay processing matrix, where $\beta = \sqrt{\frac{P_1}{\operatorname{tr}\left(\boldsymbol{U}_x(\boldsymbol{H}_{x,k}\widehat{\boldsymbol{F}}_k \widehat{\boldsymbol{F}}_k^H \boldsymbol{H}_{x,k}^H + \sigma^2 \boldsymbol{I}_{N_r})\boldsymbol{U}_x^H\right)}}$, and we then choose the relay processing and the linear receive matrices to maximize the sum-rate. The complete method to find $\boldsymbol{F}_k$, $\boldsymbol{U}_x$ and $\boldsymbol{W}_k$ is formalized by Algorithm 4.

IV. Convergence and Complexity

In terms of convergence, we use an alternating minimization approach to solve our optimization problems. At each iteration, we have a single-variable convex optimization problem; thus,

Algorithm 3 Modified Weighted-Sum Method

Define the set of weights $\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$, $\forall \alpha_i \in [0, 1]$, with $\alpha_i < \alpha_{i-1}$ and ϵ
$w_1 \leftarrow \alpha_i$
$w_2 \leftarrow 1 - \alpha_i$
solve (36)-(37) using the objective function defined by (38), adding constraints (41)-(40)
$w_1 \leftarrow 1 - \alpha_i$
$w_2 \leftarrow \alpha_i$
repeat (36)-(37)
Until $f(n) - f(n-1) \leq \epsilon$

Algorithm 4 Algorithm to find $\boldsymbol{F}_k$, $\boldsymbol{U}_x$ and $\boldsymbol{W}_k$

Choose arbitrary matrices $\boldsymbol{U}_x$ and $\boldsymbol{W}_k$, $n = 1$
fix $\boldsymbol{U}_x$ and $\boldsymbol{W}_k$ and determine $\boldsymbol{F}_k$ by solving (35)
fix $\boldsymbol{F}_k$ and jointly determine $\boldsymbol{U}_x$ and $\boldsymbol{W}_k$ using Algorithm 3
$n = n + 1$
Until $f(n) - f(n-1) \leq \epsilon$

we can find a global optimum solution of that optimization problem. This implies that after each iteration, the cost functions of our multi-variable optimization problems are non-increasing [28], [31]. This guarantees that the proposed methods will converge in terms of the cost function. However, we should mention that because our all problems are non-convex the convergence to a global minimum cannot be guaranteed.

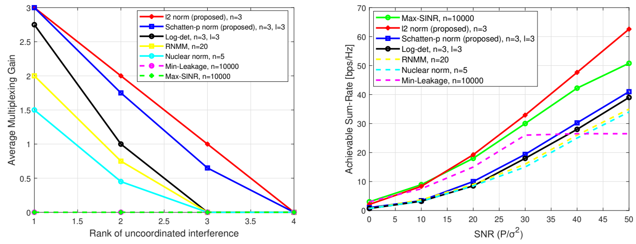
The worst case complexity of a SDP problem can be expressed as [32]:

$$\mathrm{O}(L(b\sum_{i=1}^{N} n_i^3 + b^2 \sum_{i=1}^{N} n_i^2 + b^3)) \tag{41}$$

where b and n_i are the number of equality constraints and the dimensions of the i-th semidefinite cone, respectively. Moreover, L and N are the iteration complexity of the algorithm and the number of semidefinite cone constraints, respectively. Notice that, our proposed methods are SDP-based, and their complexity can be expressed by (41). Besides, they differ from other SDP-based approaches in the literature, since by using constraint (11) we decrease the number of the inequalities of rank minimization problem compared to [4], [8]. Thus l_2 norm minimization can get lower computation time. The Schatten-p norm minimization and the algorithm in [8] have two loops (the inner loop, and the outer loop); hence, we can expect that their computation time is higher than l_1 and l_2 minimization approaches. Furthermore, the complexity of approaches in [9], [10] are $\mathrm{O}(N_t^3)$ or $\mathrm{O}(N_r^3)$ at each iteration, but they require considerably more iterations to solve their optimization problems. Therefore, it is difficult to compare the overall complexity of the methods.

V. Simulation Results

This section presents some numerical results employing the MATLAB CVX toolbox [33]. The simulations are performed over 200 channel realizations, where channel elements are independent and identically distributed (i.i.d.) Gaussian random variables with mean zero and unit variance. Moreover, we assume $\sigma^2 = 1$, while the transmit power of each user



(a) Average multiplexing gain (b) Achievable sum-rate

Fig. 2. Average multiplexing gain and achievable sum-rate for the $(10 \times 6, 4)^3$ MIMO interference channel with $P = 30$ dB and $P_1 = 0$ dB.

is $E[\|\boldsymbol{F}_k \boldsymbol{s}_k\|^2] = P$, $\forall k$, and the transmit power of each uncoordinated source is $E[\|\widetilde{\boldsymbol{F}}_{k'} \boldsymbol{x}_f\|^2] = P_{k'}$.

The methods proposed for the MIMO interference channel are compared with other IA frameworks from the literature, such as the nuclear norm minimization [4], log-det heuristic [8], the SINR maximization [9], the interference leakage minimization [10] and the re-weighted nuclear norm minimization (RNNM) [7] approaches. In addition, in the relay-aided scenario the modified weighted-sum method is compared with the leakage minimization and the weighted-sum mean squared error (WMSE) minimization from [14], a zero-forcing (ZF)-based approach from [6], and a hybrid MMSE and ZF-based scheme from [15].

A. *MIMO Interference Channel*

In this section, we consider imperfect CSI for uncoordinated interference with $\tau = 0.8$. First, we start with a $(10 \times 6, 4)^3$ MIMO interference scenario with one source of uncoordinated interference, whose transmit power is $P_1 = 0$ dB. Moreover, throughout this section we employ $\boldsymbol{\Upsilon} = \mathbf{1}_d$ for the proposed l_2 and Schatten-p norm minimization methods. Then, Fig. 2a shows the average multiplexing gain (or DoF) as a function of the rank of uncoordinated interference, which ranges from 1 to 4, when the transmit power of each user is $P = 30$ dB. As it can be noticed from the figure, the proposed schemes increase the average multiplexing gain, with the l_2 norm minimization approach achieving the best performance in this scenario. Moreover, it is worth noting that when the rank of uncoordinated interference is higher than 3, all other methods cannot provide any average multiplexing gain, while the proposed methods achieve considerably better performance.

For the same scenario, Fig. 2b and Fig.3 depict the average sum-rate (R) versus the signal-to-noise ratio (SNR) (P/σ^2) and the estimation quality (ρ) respectively, when the rank of uncoordinated interference is equal to one. As we can observe from Fig. 2b, Max-SINR [9], Min-Leakage [10] and l_2 norm minimization methods have very similar performances at low SNR, with a slight advantage for the two former in that region. Nevertheless, the l_2 norm minimization method outperforms the other schemes when $P/\sigma^2 > 20$ dB, achieving higher sum-rate than the other approaches when P increases. Such increased performance is in part due to the expansion of the feasible set of the optimization problem, which brings important gains especially when the uncoordinated interferences are

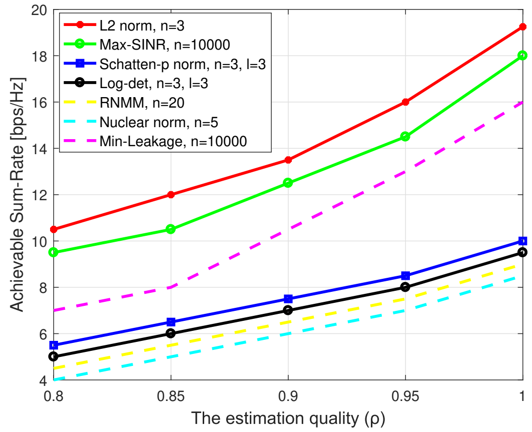


Fig. 3. The achievable sum-rate versus the estimation quality (ρ) for $(10 \times 6, 4)^3$ MIMO interference channel with one uncoordinated source, and $P = 20$ dB and $P_1 = 0$ dB

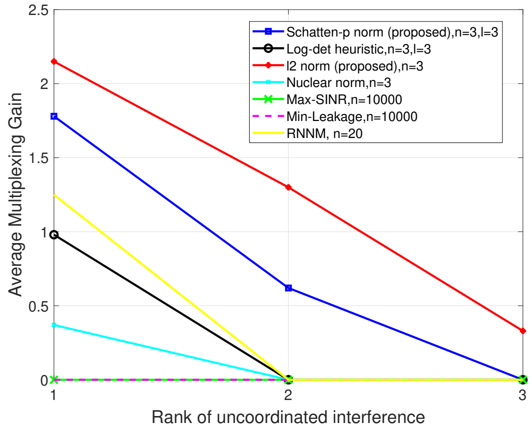


Fig. 4. Average multiplexing gain (or DoF) for the $(10 \times 6, 4)^3$ MIMO interference channel versus the rank of the uncoordinated interference sources when there are two sources of uncoordinated interference.

strong. Also, Fig.3 shows the effect of imperfect CSI, and based on it our proposed method can achieve higher sum-rate.

In Fig.4 we consider a $(10 \times 6, 4)^3$ MIMO interference system with two sources of uncoordinated interference. We plot the average multiplexing gain as a function of the rank of the first uncoordinated source, while the rank of the second uncoordinated source is fixed to 1. As it can be observed, the proposed methods noticeably outperform the other approaches. For example, when the rank of the first uncoordinated source is equal to 2, all other methods do not obtain any average multiplexing gain, while our proposed methods considerably improve the IA performance.

In order to compare the computation time of each algorithm, we show in Table I the average computation time of the proposed l_2 and Schatten-p norm methods, as well as the other benchmark algorithms for two different MIMO scenarios: $(10 \times 6, 4)^3$ and $(6 \times 4, 2)^4$. As we can observe, the average computation of the l_2 norm minimization approach is smaller than all other schemes.

Finally, we compare three different MIMO scenarios in terms of the average multiplexing gain, while we increase the number of antennas at the receiver: $(10 \times 8, 4)^3$, $(8 \times 8, 4)^3$ and $(6 \times 4, 3)^3$. Fig.5 depicts the average multiplexing gain, where we observe that both l_2 and Schatten-p norm methods achieve higher DoF, in all scenarios.

TABLE I
AVERAGE COMPUTATION TIME OF THE IA SCHEMES.

	MIMO Interference Scenario	
	$(10 \times 6, 4)^3$	$(6 \times 4, 2)^4$
l_2 norm	3.95 s	3.77 s
Schatten-p norm	23.97 s	16.02 s
Nuclear norm [4]	8.47 s	5.80 s
Min-Leakage [10]	12.15 s	16.94 s
Log-det [8]	25.45 s	14.74 s
Max-SINR [9]	13.42 s	17.50 s
RNNM [7]	21.3 s	19 s

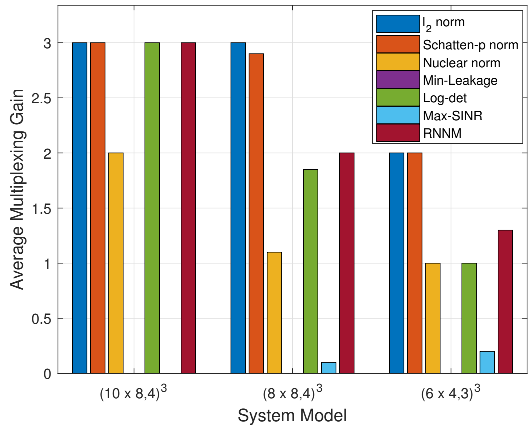
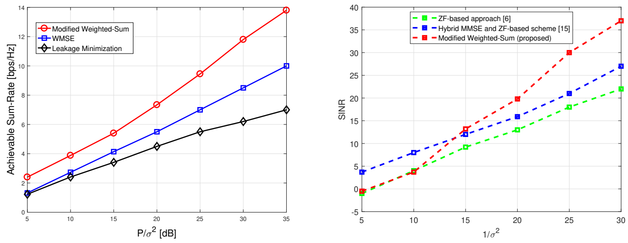


Fig. 5. Average multiplexing gain (or DoF) for three different system models.

B. Relay-Aided MIMO Interference Channel

In the relay-aided MIMO interference scenario we consider $\mu = 0.01$. Then, Fig. 6a illustrates the achievable sum-rate as a function of the SNR for $(2 \times 2, 2)^3 + 3^2$ system scenario. As we can observe, the proposed modified weighted-sum method considerably outperforms the WMSE and Leakage Minimization approaches [14]. Such improvements are obtained once our rank minimization approach tries to minimize the dimensions occupied by the interference, which increases the DoF. As a consequence, the sum-rate is also increased, especially at medium and high SNR. The Leakage Minimization approach, on the other hand, tries to minimize the energy of the interference. However, such lower-energy solutions lead to lower DoF when compared to the rank minimization approach. In addition, the WMSE method leads to an unfair distribution of the rate among the users, with some transmitters having much smaller rates than the others at medium and high SNR, jeopardizing the sum-rate.

Finally, in Fig. 6b we consider a scenario of a macro base station serving three users, and we plot the SINR as a function of $1/\sigma^2$. Then, we compare our proposed method with a zero-forcing (ZF)-based approach from [6] and a hybrid MMSE and ZF-based scheme from [15]. As we observe, our proposed



(a) Achievable sum-rate (b) Achievable SINR

Fig. 6. The average sum-rate for $(2 \times 2, 2)^3 + 3^2$ versus P/σ^2, the achievable SINR for $(4 \times 2, 2)^4 + 2^4$ relay-aided MIMO interference channel as a function of $1/\sigma^2$.

method obtains higher SINR than the other methods when the SINR is greater than 15dB, while the hybrid MMSE and ZF-based scheme performs better at low SINR.

VI. Conclusion

In this work, two rank minimization methods are proposed for the MIMO interference channel when there are sources of uncoordinated interference in the network for perfect and imperfect CSI cases. Due to the non-convexity of the rank function, we employ the l_2 norm and the Schatten-p norm to approximate the rank function. The l_2 norm is a reasonable choice when the perfect interference alignment is not attainable, while the Schatten-p norm generalizes the nuclear norm, which is usually employed in this context, but with the advantage that it provides solutions closer to the rank function when $p \to 0$. We also propose a new convex relaxation to expand the feasible set of our optimization problem, by introducing a matrix denoted by $\boldsymbol{\Upsilon}$, which provides lower rank solutions. Besides, a new method is proposed to decrease the effect of interference in the relay-aided MIMO interference channel, for which a modified weighted-sum method is proposed. The modified approach employs a set of weighting parameters in order to expand the feasible set and provide more optimal points. Finally, finding algorithms with lower complexity that require less overhead which is important for the large networks could be a topic for the future work. Also, developing fully distributed IA methods especially for the relay-aided scenarios is another interesting future topic.

Appendix A
Proof of Lemma 2

To write the Lagrange dual function, we should consider constraints (31) and (32). Moreover, according to constraint (10d), $\boldsymbol{Y}_k(1{:}d, 1{:}d) = \boldsymbol{I}_k$, where $\boldsymbol{Y}_k(1{:}d, 1{:}d)$ is a submatrix of $\boldsymbol{Y}_k$. Thus, the Lagrangian associated with (10) can be expressed as follows:

$$
\begin{aligned}
L(\boldsymbol{W}, \boldsymbol{C}, \boldsymbol{D}) = &\sum_{k=1}^{K} \mathrm{tr}\Big((\boldsymbol{S}_k - \boldsymbol{W}_k^H \boldsymbol{\Theta})\boldsymbol{D}_{1,k}\Big) + \sum_{k=1}^{K} \mathrm{tr}\Big((-\boldsymbol{Y}_k)\boldsymbol{C}_{2,k}\Big) \\
&\sum_{k=1}^{K} \mathrm{tr}(\boldsymbol{Y}_k) + \sum_{k=1}^{K} \mathrm{tr}\Big((-\boldsymbol{S}_k + \gamma \boldsymbol{I}_d)\boldsymbol{C}_{1,k}\Big) \\
&+ \sum_{k=1}^{K} \mathrm{tr}\Big(\boldsymbol{Y}_k(1{:}d, 1{:}d)\boldsymbol{D}_{3,k}\Big) - \sum_{k=1}^{K} \mathrm{tr}(\boldsymbol{D}_{3,k}\boldsymbol{I}_d) \\
&+ \sum_{k=1}^{K} \mathrm{tr}\Big((\boldsymbol{J}_k - \boldsymbol{W}_k^H \boldsymbol{H}_{k,l}\boldsymbol{V}_l)\boldsymbol{D}_{2,k}\Big), \qquad (42)
\end{aligned}
$$

where $\boldsymbol{C}_{1,k}$ and $\boldsymbol{C}_{2,k}$ are the Lagrange multipliers associated with the inequality constraints, $\boldsymbol{D}_{1,k}$, $\boldsymbol{D}_{2,k}$ and $\boldsymbol{D}_{3,k}$ are the Lagrange multipliers associated with the equality constraints, and $\boldsymbol{\Theta} = \left[\{\boldsymbol{H}_{k,l}\boldsymbol{F}_l\}_{l=1, l\neq k}^{K} \cdots \left\{\widetilde{\boldsymbol{H}}_{k,m}\widetilde{\boldsymbol{F}}_m\right\}_{m=1}^{M}\right]$. Then, by using (10d), we can rewrite the interference matrix based on $\boldsymbol{Y}_k$ as $\boldsymbol{J}_k = \boldsymbol{Y}_k(1{:}d, d+1{:}\zeta)$, recalling that $\zeta = Kd + A_m$. Furthermore, we have

$$
\mathrm{tr}\left(\boldsymbol{Y}_k \begin{pmatrix} \boldsymbol{0}_{d\times d} & \boldsymbol{0}_{d\times(\zeta-d)} \\ \boldsymbol{C}_{2,k} & \boldsymbol{0}_{(\zeta-d)\times(\zeta-d)} \end{pmatrix}\right) = \mathrm{tr}(\boldsymbol{J}_k \boldsymbol{D}_{2,k}). \qquad (43)
$$

Moreover, since $\boldsymbol{Y}_k(1{:}d, 1{:}d) = \boldsymbol{I}_d$ must be included as a constraint to the dual problem. Thus,

$$
\mathrm{tr}\left(\boldsymbol{Y}_k \begin{pmatrix} \boldsymbol{D}_{3,k} & \boldsymbol{0}_{d\times(\zeta-d)} \\ \boldsymbol{0}_{(\zeta-d)\times d} & \boldsymbol{0}_{(\zeta-d)\times(\zeta-d)} \end{pmatrix}\right) = \mathrm{tr}\Big(\boldsymbol{Y}_k(1{:}d, 1{:}d)\boldsymbol{D}_{3,k}\Big). \qquad (44)
$$

Finally, since (42) is a linear function it becomes unbounded unless the following constraints hold:

$$
\begin{aligned}
\boldsymbol{I}_\zeta - \boldsymbol{C}_{2,k} + \boldsymbol{DD}_{1,k} + \boldsymbol{DD}_{2,k} &= \boldsymbol{0}_{\zeta\times\zeta}, && (45)\\
\boldsymbol{D}_{1,k} - \boldsymbol{C}_{1,k} &= \boldsymbol{0}_{d\times d}, && (46)\\
\boldsymbol{C}_{1,k} &\succeq \boldsymbol{0}_{d\times d}, && (47)\\
\boldsymbol{C}_{2,k} &\succeq \boldsymbol{0}_{\zeta\times\zeta}, && (48)\\
\boldsymbol{H}_{k,k}\boldsymbol{F}_k\boldsymbol{D}_{1,k} + \boldsymbol{\Theta}\boldsymbol{D}_{2,k} &= \boldsymbol{0}_{N_r\times d}, && (49)
\end{aligned}
$$

where $\boldsymbol{DD}_{1,k} = \begin{pmatrix} \boldsymbol{0}_{d\times d} & \boldsymbol{0}_{d\times(\zeta-d)} \\ \boldsymbol{D}_{2,k} & \boldsymbol{0}_{(\zeta-d)\times(\zeta-d)} \end{pmatrix}$ and $\boldsymbol{DD}_{2,k} = \begin{pmatrix} \boldsymbol{D}_{3,k} & \boldsymbol{0}_{d\times(\zeta-d)} \\ \boldsymbol{0}_{(\zeta-d)\times d} & \boldsymbol{0}_{(\zeta-d)\times(\zeta-d)} \end{pmatrix}$, yielding (15) and concluding the proof.

Appendix B
Proof of Proposition 1

Using the constraint in (16) changes the Lagrangian of our optimization problem as follows:

$$\begin{aligned}L(\boldsymbol{W},\boldsymbol{C},\boldsymbol{D}) &= \sum_{k=1}^{K}\mathrm{tr}(\boldsymbol{Y}_k)+\sum_{k=1}^{K}\mathrm{tr}\Big((-\boldsymbol{S}_k+\gamma\boldsymbol{I}_d-\boldsymbol{\Upsilon})\boldsymbol{C}_{1,k}\Big)\\ &+\sum_{k=1}^{K}\mathrm{tr}\Big((\boldsymbol{S}_k-\boldsymbol{W}_k^H\boldsymbol{\Theta})\boldsymbol{D}_{1,k}\Big)-\sum_{k=1}^{K}\mathrm{tr}(\boldsymbol{D}_{3,k}\boldsymbol{I}_d)\\ &+\sum_{k=1}^{K}\mathrm{tr}\Big(\boldsymbol{Y}_k(1{:}d,1{:}d)\boldsymbol{D}_{3,k}\Big)+\sum_{k=1}^{K}\mathrm{tr}\Big((-\boldsymbol{Y}_k)\boldsymbol{C}_{2,k}\Big)\\ &+\sum_{k=1}^{K}\mathrm{tr}\Big((\boldsymbol{J}_k-\boldsymbol{W}_k^H\boldsymbol{H}_{k,l}\boldsymbol{F}_l)\boldsymbol{D}_{2,k}\Big),\end{aligned}\tag{50}$$

where it is worth noting that the matrix $\boldsymbol{\Upsilon}$ has no effect in the constraints (45)-(49) because it does not depend on any variable of the primal optimization problem. As a result, only the cost function of the dual problem may be modified, yielding (17).

Appendix C
Proof of Proposition 2

As previously mentioned, based on strong duality the optimal values of both dual and original problem are equal. Thus, let us suppose that the optimal value obtained by (15) when $\boldsymbol{\Upsilon}=0$ is P^*, while $P^*(\boldsymbol{\Upsilon})$ denotes the optimal value of (15) with a non-zero $\boldsymbol{\Upsilon}$. Moreover, let $\boldsymbol{C}_{1,k}^0$ and $\boldsymbol{D}_{3,k}^0$ be the optimal matrices of (15) when $\boldsymbol{\Upsilon}=0$, so that the maximum value of $\mathrm{tr}(\boldsymbol{C}_{1,k}-\boldsymbol{D}_{3,k})$ associated with the constraints in (15b)-(15f) is $\mathrm{tr}(\boldsymbol{C}_{1,k}^0-\boldsymbol{D}_{3,k}^0)$. On the other hand, when $\boldsymbol{\Upsilon}$ is non-zero we denote $\boldsymbol{C}_{1,k}^{\boldsymbol{\Upsilon}}$ and $\boldsymbol{D}_{3,k}^{\boldsymbol{\Upsilon}}$ as the optimal matrices. However, since the constraints of (15) remain unchanged and only the objective function is modified according to Proposition 1 we can conclude that

$$\mathrm{tr}(\boldsymbol{C}_{1,k}^0-\boldsymbol{D}_{3,k}^0)\geq\mathrm{tr}(\boldsymbol{C}_{1,k}^{\boldsymbol{\Upsilon}}-\boldsymbol{D}_{3,k}^{\boldsymbol{\Upsilon}}).\tag{51}$$

Taking each objective function into account we have

$$P^*=\mathrm{tr}(\boldsymbol{C}_1^0-\boldsymbol{D}_{3,k}^0),\tag{52}$$

$$P^*(\boldsymbol{\Upsilon})=\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}})-\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}\boldsymbol{\Upsilon})-\mathrm{tr}(\boldsymbol{D}_{3,k}^{\boldsymbol{\Upsilon}}).\tag{53}$$

Then, we can write the following

$$P^*-P^*(\boldsymbol{\Upsilon})=\mathrm{tr}(\boldsymbol{C}_1^0-\boldsymbol{D}_{3,k}^0)-\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}-\boldsymbol{D}_{3,k}^{\boldsymbol{\Upsilon}})+\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}\boldsymbol{\Upsilon}).\tag{54}$$

As long as $\mathrm{tr}(\boldsymbol{C}_1^0-\boldsymbol{D}_{3,k}^0)-\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}-\boldsymbol{D}_{3,k}^{\boldsymbol{\Upsilon}})\geq 0$ and $\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}\boldsymbol{\Upsilon})\geq 0$ we can conclude that $P^*(\boldsymbol{\Upsilon})\geq P^*-\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}\boldsymbol{\Upsilon})$, while as long as $\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}-\boldsymbol{D}_{3,K}^{\boldsymbol{\Upsilon}})-\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}\boldsymbol{\Upsilon})\leq\mathrm{tr}(\boldsymbol{C}_1^0-\boldsymbol{D}_{3,k}^0)$ we can also conclude that $P^*(\boldsymbol{\Upsilon})\leq P^*$. Therefore,

$$P^*-\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}\boldsymbol{\Upsilon})\leq P^*(\boldsymbol{\Upsilon})\leq P^*.\tag{55}$$

Thus, when $\boldsymbol{\Upsilon}$ is small $P^*(\boldsymbol{\Upsilon})$ approaches to its lower bound $(P^*-\mathrm{tr}(\boldsymbol{C}_1\boldsymbol{\Upsilon}))$. As a result, $\boldsymbol{\Upsilon}$ should be chosen as a matrix with small entries, ensuring that $\mathrm{tr}(\boldsymbol{C}_1^{\boldsymbol{\Upsilon}}\boldsymbol{\Upsilon})\geq 0$.

Appendix D
Proof of Proposition 3

The nuclear norm is defined as the sum of the singular values of the matrix, denoted by $\|.\|_*=\sum_{i=1}^{\mathrm{rank}(.)}\delta_i(.)$. Thus, let us assume that (26) is the nuclear norm of a given matrix $\boldsymbol{R}_k=\boldsymbol{G}_k^l\boldsymbol{J}_k$. Then, due to the fact that the SVD decomposition of $\boldsymbol{J}_k$ equals to $\boldsymbol{\Lambda}_k\boldsymbol{\Pi}_k\boldsymbol{\Psi}_k^H$, $\boldsymbol{G}_k^l$ must be equal to $\boldsymbol{\Lambda}_k\boldsymbol{\Phi}_k^l\boldsymbol{\Lambda}_k^H$ in order to (26) and (27a) be equivalent expressions, where $\boldsymbol{\Phi}_k^l\in\mathbb{R}^{d\times d}$ is given by

$$\boldsymbol{\Phi}_k^l=\begin{pmatrix}\frac{p}{\left(\delta_1(J_k^l)+\xi\right)^{1-p}} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{p}{\left(\delta_d(J_k^l)+\xi\right)^{1-p}}\end{pmatrix}.\tag{56}$$

Therefore, we have that

$$\|\boldsymbol{R}_k\|_*=\left\|\boldsymbol{\Lambda}_k\boldsymbol{\Phi}_k^l\boldsymbol{\Lambda}_k^H\boldsymbol{\Lambda}_k\boldsymbol{\Pi}_k\boldsymbol{\Psi}_k^H\right\|_*=\left\|\boldsymbol{\Lambda}_k\boldsymbol{\Omega}_k\boldsymbol{\Psi}_k^H\right\|_*,\tag{57}$$

where $\boldsymbol{\Omega}_k=\boldsymbol{\Phi}_k^l\boldsymbol{\Pi}_k$ is a diagonal matrix. Finally, according to (57) we conclude that $\|\boldsymbol{G}_k^l\boldsymbol{J}_k\|_*=\sum\limits_{i=1}^{d}\frac{p\,\delta_i(\boldsymbol{J}_k)}{\left(\delta_i(\boldsymbol{J}_k^l)+\xi\right)^{1-p}}$.

References

[1] G. Liu, M. Sheng, X. Wang, W. Jiao, Y. Li, and J. Li, "Interference alignment for partially connected downlink MIMO heterogeneous networks," *IEEE Trans. Commun.*, vol. 63, pp. 551–564, Jan 2015.

[2] G. Bresler, D. Cartwright, and D. Tse, "Feasibility of interference alignment for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 60, pp. 5573–5586, Jul 2014.

[3] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3425–3441, Jul 2008.

[4] D. S. Papailiopoulos and A. G. Dimakis, "Interference alignment as a rank constrained rank minimization," *IEEE Trans. Sig. Process.*, vol. 60, pp. 4278–4288, Aug 2012.

[5] T. Gou and S. A. Jafar, "Degrees of freedom of the K user M × N MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 56, pp. 6040–6057, Dec 2010.

[6] X. Chen, S. Song, and K. Letaief, "Interference alignment in dual-hop MIMO interference channel," *IEEE Trans. Wirel Commun.*, vol. 13, pp. 1274–1283, Mar 2014.

[7] G. Sridharan and W. Yu, "Linear beamformer design for interference alignment via rank minimization," *IEEE Trans. Sig. Process.*, vol. 63, pp. 5910–5923, Nov 2015.

[8] H. Du, T. Ratnarajah, M. Sellathurai, and C. B. Papadias, "Reweighted nuclear norm approach for interference alignment," *IEEE Trans. Commun.*, vol. 61, pp. 3754–3765, Sep 2013.

[9] S. W. Peters and R. W. Heath Jr, "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, pp. 206–218, Oct 2011.

[10] S. W. Peters and R. W. Heath Jr, "Interference alignment via alternating minimization," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2445–2448, IEEE, 2009.

[11] S. Mollaebrahim, P. Ghari, M. Fazel, and M. Imran, "Designing precoding and receive matrices for interference alignment in mimo interference channels," in *IEEE Global Communications Conference (Globecom)*, IEEE, 2017.

[12] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, pp. 10–21, Jun 2011.

[13] J. Zhang, Z. Ming Ho, E. A. Jorswieck, and M. Haardt, "Interference neutralization for non-regenerativetwo-way relaying networks—part ii: Optimization," *IEEE Trans. Sig. Process.*, vol. 63, pp. 4557–4571, Jun 2015.

[14] K. T. Truong, P. Sartori, and R. W. Heath, "Cooperative algorithms for mimo amplify-and-forward relay networks," *IEEE Trans. Sig. Process.*, vol. 61, pp. 1272–1287, March 2013.

[15] C. Le, S. Moghaddamnia, and J. Peissig, "A hybrid optimization approach for interference alignment in multi-user mimo relay networks under different csi," *IEEE Trans. Wirel Commun.*, vol. 16, pp. 7834–7847, Dec 2017.

[16] D. Papsdorf, Y. Silva, and A. Klein, "Multicast interference alignment in a multi-group multi-way relaying network," in *IEEE International Conference on Communications (ICC)*, IEEE, 2017.

[17] Q. Shi, M. Razaviyayn, Z. Luo, and C.He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Sig. Process.*, vol. 59, pp. 4331–4340, Sep 2011.

[18] S. Berger, M. Kuhn, A. Wittneben, T. Unger, and A. Klein, "Recent advances in amplify-and-forward two-hop relaying," *IEEE Commun. Mag.*, vol. 47, pp. 50–56, July 2009.

[19] O. El Ayach, S. W. Peters, and R. W. Heath Jr, "The practical challenges of interference alignment," *IEEE Wireless Commun.*, vol. 20, pp. 35–42, Feb 2013.

[20] F. Quitin, M. Rahman, R. Mudumbai, and U. Madhow, "Distributed beamforming with software-defined radios: Frequency synchronization and digital feedback," in *IEEE Global Communications Conference (Globecom)*, IEEE, 2012.

[21] R. Rogalin, O. Bursalioglu, H. Papadopoulos, and G. Caire, "Scalable synchronization and reciprocity calibration for distributed multiuser mimo," *IEEE Wireless Commun.*, vol. 13, pp. 1815–1831, Apr 2014.

[22] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Sig. Process.*, vol. 58, pp. 4771–4782, May 2010.

[23] L. Ruan, V. K. Lau, and M. Z. Win, "The feasibility conditions for interference alignment in MIMO networks," *IEEE Trans. Sig. Process.*, vol. 61, pp. 2066–2077, Jan 2013.

[24] M. Fazel, *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

[25] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[26] G. Golub and C. Van Loan, *Matrix computations*. Johns Hopkins University Press, 1996.

[27] S. Boyd, L. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994.

[28] D. P. Bertsekas, *Nonlinear Programming*. Singapore. Athena Scientific, 1999.

[29] D. Ge, X. Jiang, and Y. Ye, "A note on the complexity of l p minimization," *Mathematical programming*, vol. 129, pp. 285–299, Oct 2011.

[30] I. Y. Kim and O. L. D. weck, "Adaptive weighted-sum method for bi-objective optimization: pareto front generation," *Structural and Multidisciplinary Optimization*, vol. 29, pp. 631–657, Sep 2005.

[31] J. Bezdek and R. Hathaway, "Convergence of alternating optimization," *Neural, Parallel and Scientific Computations.*, vol. 11, pp. 351–368, Dec 2003.

[32] I. Polik and T. Terlaky, "Interior point methods for nonlinear optimization," *Nonlinear Optimization.*, Jan 2010.

[33] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," 2008.