Backhaul Limited Asymmetric Cooperation for MIMO Cellular Networks via Semidefinite Relaxation

Fuxin Zhuang, Student Member, IEEE, Vincent K.N. Lau, Fellow, IEEE.

Abstract—Multicell cooperation has recently attracted tremendous attention because of its ability to eliminate intercell interference and increase spectral efficiency. However, the enormous amount of information being exchanged, including channel state information and user data, over backhaul links may deteriorate the network performance in a realistic system. This paper adopts a backhaul cost metric that considers the number of active directional cooperation links, which gives a first order measurement of the backhaul loading required in asymmetric Multiple-Input Multiple-Output (MIMO) cooperation. We focus on a downlink scenario for multi-antenna base stations and single-antenna mobile stations. The design problem is minimizing the number of active directional cooperation links and jointly optimizing the beamforming vectors among the cooperative BSs subject to signal-to-interference-and-noise-ratio (SINR) constraints at the mobile station. This problem is non-convex and solving it requires combinatorial search. A practical algorithm based on smooth approximation and semidefinite relaxation is proposed to solve the combinatorial problem efficiently. We show that semidefinite relaxation is tight with probability 1 in our algorithm and stationary convergence is guaranteed. Simulation results show the saving of backhaul cost and power consumption is notable compared with several baseline schemes and its effectiveness is demonstrated.

Index Terms—Cooperation, cellular networks, multicell, interference, combinatorial, beamforming, semidefinite relaxation

I. INTRODUCTION

Multicell cooperation (MCP) is a promising technique, which can be used to dramatically improve the performance of cellular networks by coordinating the base stations (BSs) through high speed backhaul links [1]. User data and channel state information (CSI) can be shared over the backhaul links to jointly encode and transmit data signals to users with multicell cooperative processing in order to exploit the intercell interference, which is known to be a limiting factor in conventional cellular networks.

Despite the great potential in combating intercell interference, the enormous signaling overhead incurred during the information exchange in the backhaul links is quite unrealistic in a practical system, where the backhaul links have finite capacity [2], [3]. There are several existing approaches in the literature to cope with this limitation. Interference coordination is introduced to combat intercell interference by exchanging CSI (instead of data payload), and hence, it can significantly reduce the backhaul loading. For example, multicell joint scheduling, power control and beamforming have been studied extensively in the literature [4], [5]. These optimization problems are well-known to be difficult because there is no known convex transformation available due to the inherit non-convexity of the SINR expression. However, a particular formulation which minimizes the total transmit power subject to the SINR constraints on single-antenna users over frequency flat channels is known to have an efficient global optimal solution by exploring semi-definite relaxation (SDR), second-order cone programming and uplink-downlink duality theory [1], [6]–[8]. There are also some works that study coordinated beamforming designs with distributed implementation in multicell network. For instance, in [7], coordinated beamforming with limited intercell information exchange is considered but the solution only applies to the sum-power minimization with SINR constraints. In [9], a distributed algorithm is designed to achieve the max-min rate fairness point on the Pareto boundary by means of uplink-downlink duality.

On the other hand, cooperative MIMO is a technique which has superior performance compared with the coordinated beamforming schemes at the expense of a huge backhaul loading requirement. There is much literature devoted to reducing the backhaul loading in cooperative MIMO schemes without sacrificing the merit of interference exploitation. For example, clustered multi-cell cooperation [10]–[12] has been proposed to reduce the backhaul loading by limiting the cooperation size. The clustering of BSs can be either static [11] or dynamic [12]. While the intra-cluster interference can be effectively mitigated, the system capacity is still limited by the inter-cluster interference [1]. Recently, distributed multicell processing schemes with data sharing are also developed in [13], [14]. In [13], the distributed downlink beamforming design is recasted into a linear minimum mean-square-error (LMMSE) estimation problem. [14] devised a distributed design with close-to-optimal performance using only local CSI. While these designs achieve relatively good performance gains, the huge backhaul cost due to data sharing between the BSs still remains.

Another example is called opportunistic cooperation which dynamically engages the system in different cooperation modes with various backhaul requirements, depending on the network channel conditions. For example, [1] mentioned a few cooperation modes, such as interference coordination, full cooperation, rate-limited cooperation and relay-assisted cooperation.

In this paper, we consider the minimization of the backhaul...
cost associated with MIMO cooperation subject to the SINR constraints of all the mobiles in the downlink of multi-cell networks. Unlike conventional works on MIMO cooperation, we consider asymmetric cooperation between base stations so as to allow more flexibility to reduce the backhaul requirement. An example of asymmetric cooperation between base stations is illustrated in Figure 1, where BS-1 shares the data stream \( s_1 \) with BS-2 so as to cooperatively exploit the strong interference link to MS-1. BS-2 does not share the data stream in return because the interference link from BS-1 to MS-2 is weak. Directional MIMO cooperation has been considered in [15]. A heuristic algorithm is proposed to dynamically select the directional cooperation links under a finite-capacity backhaul, subject to the evaluation of benefits and costs. However, the simple zero-forcing (ZF) transmit beamforming design assumed in [15] in the directional cooperation cannot completely eliminate the undesired interference as in the full cooperation case. This is because the directional cooperation will result in data propagation which cannot be handled by a simple ZF scheme, and the true design is highly non-trivial in a directional cooperation topology. There are various technical challenges associated with the minimization of backhaul loading subject to SINR constraints in MIMO cooperation. They are elaborated as follows:

1) **Appropriate Metric of Backhaul Cost:** To quantify the backhaul cost, one obvious metric of backhaul loading is the average b/s/Hz in the backhaul to support the data streams to the mobiles. However, such a metric reveals too much detail, and the associated problem is highly combinatorial [2]. In this paper, we consider another metric, namely the number of active directional cooperation links. Since the number of active directional cooperation links is proportional to the backhaul loading (number of data streams) in the MIMO cellular network, this metric gives a first order measurement of the backhaul loading (in degrees of freedom) required to support the asymmetric MIMO cooperation without revealing too much unnecessary detail.

2) **Combinatorial Optimization due to Backhaul Cost:** The proposed backhaul metric involves counting the number of active cooperation links from the CP to the BSs, and hence, the associated optimization problem is combinatorial in nature. In fact, the proposed metric can be expressed as the mixed \( l_0/l_2 \) norm of the beamforming vector at each BS. Hence, the associated problem becomes an \( l_0/l_2 \) norm minimization problems, which is NP-hard in general.

3) **Non-convexity due to SINR Constraints:** Another difficulty of the optimization problem is the SINR constraints (requirement) of all the mobiles in the systems. These SINR constraints are non-convex due to the interference coupling between the users. As such, this makes the \( l_0/l_2 \) norm minimization problem different from the standard form of compressed sensing recovery [16], [17], and therefore, these standard solutions cannot be directly applied.

In this paper, we tackle the above challenges of \( l_0/l_2 \) norm optimization under non-convex SINR constraints using the smooth approximation method [18], [19] and SDR [6], [20]. Specifically, we show that the original \( l_0/l_2 \) norm minimization problem is asymptotically equivalent to a sequence of smooth minimization problems. We derive low complexity solutions by the SDR of the non-convex SINR constraints and show that the solution of the relaxed problem is always tight with probability 1. Convergence of the low complexity algorithm to the stationary point of the original non-smooth problem is guaranteed. It is worth noting that in a very recent paper [21], the authors consider a similar problem scenario but with substantially different formulation and solution approach. A reweighted \( l_1 \) norm minimization method and a heuristic iterative link removal algorithm are proposed in [21] but these suboptimal schemes still suffer from a significant performance loss. The simulation result shows that our proposed algorithm can effectively reduce the backhaul loading required in the asymmetric MIMO cooperation compared with various baselines.

This paper is organized as follows. In Section II, we introduce the system model and formulate the combinatorial optimization problem, which minimizes the cooperation links subject to SINR constraints. In Section III, we use the smooth functions to approximate the \( l_0/l_2 \) norm and apply SDR to the non-convex quadratic constraints to obtain the approximation problem. In Section IV, a low complexity algorithm is proposed to solve the combinatorial problem. Based on this, we show that the SDR is always tight with probability 1. Section V investigates the performance of the algorithm by simulation results. Section VI draws some concluding remarks.

**Notation:** We adopt the notation of using boldface for vectors \( a \) (lower case) and matrices \( A \) (upper case). The transpose operator and the complex conjugate transpose operator are denoted by the symbols \( (\cdot)^T \) and \( (\cdot)^H \), respectively. Tr(\( \cdot \))

---

1. In [6], the authors established the strong duality results and the tightness of homogeneous SDR. In our case, the problem is inhomogeneous and we have extended the proof for strong duality under sufficient conditions.
is the trace of the square matrix argument. \( \mathbf{I} \) and \( \mathbf{0} \) denote, respectively, the identity matrix and the matrix with zero entries (their size is determined from context). \( \mathbf{I} \) represents a vector of 1s and its size depends on the context. The kronecker product is denoted as \( \otimes \). For any complex vector \( \mathbf{x} \), we use \( \| \mathbf{x} \|, \| \mathbf{x} \|_0, \| \mathbf{x} \|_2 \) to represent the Euclidean norm, \( l_0 \) norm and \( l_2 \) norm of \( \mathbf{x} \). The curled inequality symbol \( \succeq \) (its strict form \( > \) and reverse form \( \prec \)) is used to denote generalized inequality: \( \mathbf{A} \succeq \mathbf{B} \) means that \( \mathbf{A} - \mathbf{B} \) is a Hermitian positive semidefinite matrix (\( \mathbf{A} > \mathbf{B} \) for positive definiteness and \( \mathbf{A} \preceq \mathbf{B} \) for negative semidefiniteness). The \( \dagger \) symbol is used to denote pseudo-inverse: \( \mathbf{A}^{\dagger} \) means the pseudo-inverse of \( \mathbf{A} \). \( \mathcal{N} \) and \( \mathcal{U} \) denotes Gaussian and uniform distribution respectively.

## II. System Model and Preliminaries

### A. Asymmetric MIMO Cooperation

Consider an MCP wireless system where several base stations jointly serve a number of single-antenna mobile stations (MSs) over a common frequency band, as shown in Figure 2. Let the number of base stations be \( N \), each equipped with \( L \) transmit antennas, and let the total number of users in the system be \( K \). We consider a centralized MCP backhaul model with a central processor (CP). All the user data and CSI are available at the CP for joint processing. Each BS is connected to the CP via backhaul links. The CSI is assumed to be perfect in this paper.

We consider asymmetric MIMO cooperation among the \( N \) BSs in the MCP network to mitigate interference. Denote the set of BSs that have acquired the data signal for the \( k \)-th user as \( \mathcal{Q}_k \), where \( \mathcal{Q}_k \subseteq \{1, 2, \ldots, N\} \). Specifically, cooperation is asymmetric, and the \( \mathcal{Q}_k \) could be any subset of \( \{1, 2, \ldots, N\} \). For example, in Figure 2(c), the cooperation is asymmetric since \( \mathcal{Q}_1 = \{1, 2\}, \mathcal{Q}_2 = \{2, 3\}, \mathcal{Q}_3 = \{1, 3\} \) and \( \mathcal{Q}_4 = \{3, 4\} \). Therefore, the \( n \)-th BS that is in \( \mathcal{Q}_k \) can increase the received SINR of the \( k \)-th user by exploiting the link from the \( n \)-th BS to the \( k \)-th user, but it also incurs the backhaul cost of transmitting the data signal from the CP over the backhaul. Conventional full Cooperative-MIMO (CoMIMO) and clustered Co-MIMO schemes are special cases of the above asymmetric MIMO cooperation, as illustrated in Figure 2(a) and 2(b). In general, full symmetric MIMO cooperation is not always necessary, as previously illustrated in Figure 1. Using the proposed asymmetric MIMO cooperation, flexible cooperation patterns can be engaged to satisfy the SINR requirements of users with minimum backhaul cost.

Let \( s_k \in \mathbb{C} \) represents the information signal for the \( k \)-th user with unit energy. Let \( \mathbf{w}_{n,k} \in \mathbb{C}^L, \forall n = 1, \ldots, N \), be the beamforming vector used by the \( n \)-th BS to pre-code the information signal for the \( k \)-th user, and \( \mathbf{w}_k = [\mathbf{w}_{1,k}^T, \mathbf{w}_{2,k}^T, \ldots, \mathbf{w}_{N,k}^T]^T \in \mathbb{C}^{LN} \) is the aggregate beamforming vector of the \( N \) BSs for the \( k \)-th user. Note that \( \mathbf{w}_{n,k} = \mathbf{0} \) if there is no data signal of the \( k \)-th user at the \( n \)-th BS. Let \( \mathbf{h}_{n,k} \in \mathbb{C}^L \) denote the complex channel fading between the \( n \)-th BS and the \( k \)-th user. The vector \( \mathbf{h}_k = [\mathbf{h}_{1,k}^T, \mathbf{h}_{2,k}^T, \ldots, \mathbf{h}_{N,k}^T]^T \in \mathbb{C}^{LN} \) is the aggregate channel fading from all the \( N \) BSs to the \( k \)-th user, which is modeled as \( \mathcal{CN}(0, I) \). The channel is assumed to be static over the backhaul.

Fig. 2. Three multicell cooperation scenarios: (a) Cluster MIMO with two BSs forming a cooperation cluster. (b) Conventional full cooperation MIMO with each BS sharing information with all the other BSs. (c) Asymmetric MIMO topology with partial directional cooperation links activated.
transmission period. The received signal at the \( k \)-th user is given by

\[
y_k = \frac{h_k^H w_k s_k + \sum_{m \neq k} h_k^H w_m s_m + n_k}{\sigma_k^2}, \quad \forall k
\]  

(1)

where \( n_k \) is the complex Gaussian white noise with zero mean and variance \( \sigma_k^2 \). The SINR at the \( k \)-th user is expressed as

\[
\text{SINR}_k \triangleq \frac{|h_k^H w_k|^2}{\sum_{m \neq k} |h_k^H w_m|^2 + \sigma_k^2}, \quad \forall k = 1, ..., K. 
\]

(2)

B. Conventional Optimal Beamforming Problem with Full Cooperation

If there is no backhaul cost, all the BSs should participate in cooperative MIMO processing and they should mutually share all their data streams over the network. This corresponds to the following design problem that minimizes the total power consumption subject to the SINR requirements of each user:

\[
\text{(OBP)} \quad \min_{\{w_k\}_{k=1}^K} \sum_{k=1}^K \|w_k\|_2^2 \\
\text{s.t.} \quad \sum_{m \neq k} |h_k^H w_m|^2 + \sigma_k^2 \geq \gamma_k, \forall k. 
\]

(3)

This optimal beamforming problem can be reformulated into a separable homogeneous Quadratically-Constrained-Quadratic-Program (QCQP), which is a non-convex problem. Furthermore, it was shown in \([22]\) that a separable homogeneous QCQP problem is NP-hard in general. Nonetheless, it is elegantly solved by \([6]\) using the Perron-Frobenius theory for matrices with nonnegative entries. Generalized uplink-downlink duality using the Lagrangian theory is introduced in \([7]\) to efficiently solve the above (OBP). It is also worth noting that when the channel vectors are confined to the far-field, line-of-sight scenario (i.e., Vandermonde channel), the OBP can be reformulated as a convex problem \([23]\).

However, in reality, the full sharing of user data is costly and highly impractical. It is possible and desirable to meet the same SINR constraints for each user, while requiring partial sharing of user data. This is a challenging combinatorial beamforming design problem, and it will be discussed in the following subsection \( D \).

C. Optimal Beamforming with Asymmetric Cooperation

For a given asymmetric cooperation topology, which means all the \( Q_a \) are fixed in the MCP network with partially shared data signals, partial entries in the beamforming weights \( w_k \) should be forced to zero. Let \( m_k = [m_{1,k}, m_{2,k}, ..., m_{N,k}] \in \mathbb{C}^{1 \times N} \), where \( m_{n,k} \in \{0,1\} \). Specifically, \( m_{n,k} = 0 \) means that the \( n \)-th BS has obtained the information signal for the \( k \)-th user while \( m_{n,k} = 1 \) means the opposite. Thus, we have

\[
\|m_k\|_0 \leq N - 1
\]

(4)

since at least one BS is used to transmit signal to the \( k \)-th user. Let \( \hat{m}_k = m_k \otimes 1_{1 \times L} \) and \( M_k = \text{diag}(\hat{m}_k) \). Therefore, we have

\[
M_k \succeq 0 \quad \text{and} \quad \text{rank}(M_k) \leq LN - L.
\]

(5)

Like in the OBP with full cooperation case, the design problem is to minimize the total power subject to the QoS requirements of each user, which gives the following optimization problem:

\[
\text{(OBP-AC)} \quad \min_{\{w_k\}_{k=1}^K} \sum_{k=1}^K \|w_k\|_2^2 \\
\text{s.t.} \quad \sum_{m \neq k} |h_k^H w_m|^2 + \sigma_k^2 \geq \gamma_k, \forall k;
\]

\[
M_k w_k = 0, \forall k.
\]

(6)

(7)

where the constraint (7) is to force the corresponding entries in \( w_k \) to be zero. Note that \( M_k \) are different matrices, which means different \( w_k \) may have different deactivated entries.

Thus this optimization problem cannot be degenerated to (OBP) by simply omitting the deactivated entries and reducing the size of \( w_k \). This separable homogeneous QCQP problem with additional equality constraints (individual shaping constraints as termed in \([20]\)) is still NP-hard. However, we manage to prove that the SDR is tight with probability 1, which will be discussed in detail in Section IV.B.

D. Optimal Beamforming with Backhaul Cost

We first define the backhaul cost associated with the asymmetric MIMO cooperation described in the subsection \( A \). For a given asymmetric MIMO cooperation scheme \( \{w_1, w_2, ..., w_K\} \), the backhaul cost is defined as the degree of freedom (number of data streams) over the backhauls from the CP to the \( N \) BSs.

When \( w_{n,k} \neq 0 \), it means that the data signal for the \( k \)-th user has to be shared from the CP to the \( n \)-th BS and this would consume backhaul for one unit using the cost metric we defined above. Correspondingly, when \( w_{n,k} = 0 \), the user data for the \( k \)-th user will not be shared from the CP to the \( n \)-th BS, and hence, it will not consume the backhaul. This specific structure design on the beamforming vectors has recently drawn a lot of attention (see \([24]\) for backhaul reduction and \([25]\) for antenna selection) and it is usually solved via the \( l_0/l_q \) norm\(^2\) approach. In this paper, we consider the common setup with \( q = 2 \). \([24],[26]\). Let \( \tilde{w}_k = [\|w_{1,k}\|, \|w_{2,k}\|, \ldots, \|w_{N,K}\|]^T \in \mathbb{R}^N \). Then \( \|\tilde{w}_k\|_{0,2} = \|\tilde{w}_k\|_0 \). Since at least one BS should acquire the data signal \( s_k \) from the CP to serve the \( k \)-th user, one of the backhaul usages of the \( k \)-th user should not be counted as additional backhaul cost. Thus, the backhaul cooperation cost metric is given below:

\[
C_B = \sum_{k=1}^K \|w_k\|_{0,2} - K,
\]

(8)

where the term \( K \) is to subtract the associated BSs from the cooperating backhaul cost. Specifically, we consider the

\(^2\)Note any \( l_0/l_q \) norm is a group sparse norm with \( q \geq 0 \).
following optimization problem:

\[
\begin{align*}
\text{(P0)} \quad \min_{\{w_k\}_{k=1}^K} & \quad \sum_{k=1}^K \|w_k\|_{0,2} + \epsilon \sum_{k=1}^K \|w_k\|_2^2 \\
\text{s.t.} & \quad \sum_{m \neq k} |h_k^H h_m|^2 + \sigma_k^2 \geq \gamma_k, \forall k,
\end{align*}
\]

where \( \epsilon > 0 \) is the power cost with respect to the total transmit power in the cellular network. This power cost term can be tuned to control the tradeoff between the backhaul cooperation cost and power consumption. An interesting relation between (OBP) and (P0) is that the \( l_2 \) norm is used for the power minimization in (P0) is actually an early approach to approximate the sparsest solution and is called the method of frames [27]. Thus, the solution of (OBP) actually gives a rough estimate for the problem (P0).

In this paper, we are going to solve (P0) and the key challenges are due to the combinatorial \( l_0/l_2 \) norm minimization, which is NP-hard, and the non-convex homogeneous quadratic constraints. The non-convex SINR constraints turn (P0) into a non-standard \( l_0/l_2 \) norm minimization problem that is different from the traditional compressed sensing problem, where linear equality constraints are assumed. In [25], an \( l_1/l_\infty \) norm squared norm approach is proposed to approximate the \( l_0/l_\infty \) norm and the problem is solved using SDP transformations. However, this approach cannot be directly extended to the more general multi-antenna MCP scenario we consider in this paper, because the approach in [25] relies on coupling across all the beamforming vectors due to the group norm. Thus, so far, there is no reliable and general method to deal with \( l_0/l_2 \) (or any other \( l_0/l_q \) norm minimization with nonlinear and non-convex constraints. In this paper, we shall first introduce an approximation problem of (P0) and then deploy the SDR technique to solve the approximation problem. Later we prove that the approximation problem and the original problem (P0) are asymptotically equivalent.

III. PROBLEM APPROXIMATION AND TRANSFORMATION

A. Transformation 1 - Approximating the \( l_0 \) Norm with Smooth Functions

The problem of using the \( l_0/l_2 \) norm for a combinatorial search for its minimization is due to the fact that the \( l_0 \) norm is a non-convex and non-smooth function. [18] and [19] state an idea to approximate this non-smooth \( l_0 \) by a continuously differentiable function, and minimize it by means of calculus (e.g., steepest descent method). The advantage of smoothing methods is that we solve optimization problems with continuously differentiable functions for which there are rich theory and powerful solution methods [29]. A local minimizer or stationary point of the original non-smooth problem can be guaranteed to be found by updating the smooth parameter [18]. The efficiency of smoothing methods will depend on the smooth approximation function, the solution method for the smooth optimization problem and the updating scheme for the smoothing parameter, which is \( \theta \) in our case.

In this paper, we consider the following family of smooth functions with the smoothing parameter \( \theta \) [19]:

\[
f_\theta(s) \triangleq \exp\left(-\frac{s^2}{2\theta^2}\right).
\]

It is easily observed that

\[
\lim_{\theta \to 0} f_\theta(s) = \begin{cases} 
1, & \text{if } s = 0 \\
0, & \text{if } s \neq 0.
\end{cases}
\]

Then, by defining

\[
F_\theta(w_k) = \sum_{i=1}^N f_\theta(\|w_i, k\|_2),
\]

and

\[
F_\theta(w_1, \ldots, w_K) = \sum_{k=1}^K F_\theta(w_k),
\]

it is readily verified that

\[
\sum_{k=1}^K \|w_k\|_{0,2} \approx KN - F_\theta(w_1, \ldots, w_K)
\]

for small values of \( \theta \), and the approximation is tight when \( \theta \to 0 \). Consequently, the minimum \( l_0/l_2 \) norm solution can be found by maximizing \( F_\theta(w_1, \ldots, w_K) \) (subject to the SINR constraints) for a sufficiently small value of \( \theta \). Note that the value of \( \theta \) determines the smoothness of the function \( F_\theta \). The larger the value of \( \theta \), the smoother \( F_\theta \), but the worse the approximation to the \( l_0/l_2 \) norm. On the other hand, the smaller the value of \( \theta \), the closer the behavior of \( F_\theta \) to the \( l_0/l_2 \) norm. This smooth maximization problem can be solved using a projected gradient algorithm. [19] and [18] propose an idea to use a decreasing sequence for the smooth parameter \( \theta \) to gradually approach the actual maximum for smaller values of \( \theta \).

Replacing the \( l_0/l_2 \) norm objective with \( F_\theta \), we have

\[
\begin{align*}
\text{(P1}(\theta) \quad \max_{\{w_k\}_{k=1}^K} & \quad F_\theta(w_1, \ldots, w_K) - \epsilon \sum_{k=1}^K \|w_k\|_2^2 \\
\text{s.t.} & \quad \sum_{m \neq k} |h_k^H h_m|^2 + \sigma_k^2 \geq \gamma_k, \forall k.
\end{align*}
\]

Let \( x = [w_1^T w_2^T \ldots w_K^T]^T \in \mathbb{C}^{KLN} \) and \( W \) (at the top of the next page) denote the feasible set of problems (P1(\theta)) and (P0). Let \( p(x) = \|x\|_{0,2} + \epsilon \|x\|_2^2 = \sum_{k=1}^K \|w_k\|_{0,2} + \epsilon \sum_{k=1}^K \|w_k\|_2^2 \), then (P0) can be rewritten as

\[
\min_{x \in W} p(x).
\]

We assume (P0) is always feasible and let \( P^* \) be the set of such optimal solutions and \( g(x, \theta) = -F_\theta(x) + \epsilon \|x\|_2^2 \), then (P1(\theta)) is the same as

\[
\min_{x \in W} g(x, \theta).
\]

Assumption 1. There exists a finite set \( S^* \subset \mathbb{C}^{KLN} \) having the property that, for any \( \theta \in \Theta \subseteq R \), a point \( x(\theta) \in S^* \) exists such that

\[
x(\theta) \in \arg\min_{x \in W} g(x, \theta).
\]
\[\mathcal{W} = \left\{ x = [w_1^T \ w_2^T \ ... \ w_K^T]^T \mid \text{SINR}_k = \frac{|h_k^H w_k|^2}{\sum_{m \neq k} |h_k^H w_m|^2 + \sigma_k^2} \geq \gamma_k, \ \forall k = 1, ..., K \right\}\] (16)

By Assumption 1, we assume that the feasible set \(\mathcal{W}\) is always nonempty and there exists an optimal solution for problem (P1(\(\theta\))) for any \(\theta\) value.

**Theorem 1 (Asymptotic Equivalence between (P1(\(\theta\))) and (P0)).** (Theorem 3.2.1, [30]) Let \(\{\theta^j\} \subset \Theta\) be an infinite sequence such that

\[
\lim_{j \to \infty} g(x, \theta^j) = p(x), \ \forall x \in \mathcal{W}. \quad (20)
\]

Under Assumption 1, there exists a finite index \(j^*\) such that, for any \(j \geq j^*\), problem (P1(\(\theta^j\))), with \(\theta = \theta^j\), has a solution \(x^j\) that also solves the original problem (P0).

**Definition 1: (Stationary point of non-smooth problem (P0)).** Stationary point of the non-smooth problem (P0) is defined as \(x^j\) which is shown as follows:

\[w_j^* \triangleq w_j w_j^H, \ \forall j = 1, ..., J\] (21)

Then (P1(\(\theta^j\))) can be written as:

\[
\max_{\{w_k, w_j\}_{k=1}^K} F_\theta(w_1, ..., w_K) - \epsilon \sum_{k=1}^K \|w_k\|_2^2
\]

s.t. \(\text{Tr}(H_k W_k) - \gamma_k \sum_{m \neq k} \text{Tr}(H_k W_m) - \gamma_k \sigma_k^2 \geq 0\)

\[W_k = w_k w_k^H, \ \forall k.\] (22)

This reformulated problem still has a non-convex constraint \(W_k = w_k w_k^H\). Now observe the following equivalence:

\[W_k = w_k w_k^H \iff W_k \succeq w_k w_k^H, \ \text{rank}(W_k) \leq 1.\]

\[\iff W_k \succeq 0, \ \text{rank}(W_k) \leq 1.\] (23)

The constraint \(W_k \succeq w_k w_k^H\) is convex and can be formulated as a Schur complement [31]:

\[
\begin{bmatrix}
W_k & w_k \\
\w^H_k & 1
\end{bmatrix} \succeq 0.
\] (24)

In SDR approximation, the rank constraint on \(W_k\) is dropped to obtain a relaxed problem:

\[
(P2(\theta))\]

\[
\max_{\{w_k, w_j\}_{k=1}^K} F_\theta(w_1, ..., w_K) - \epsilon \sum_{k=1}^K \|w_k\|_2^2
\]

s.t. \(\text{Tr}(H_k W_k) - \gamma_k \sum_{m \neq k} \text{Tr}(H_k W_m) - \gamma_k \sigma_k^2 \geq 0\)

\[
\begin{bmatrix}
W_k & w_k \\
\w^H_k & 1
\end{bmatrix} \succeq 0, \ \forall k.
\] (25)

Thus, (P2(\(\theta\))) is a smooth optimization problem. As mentioned above, an optimization problem with continuously differentiable functions has prosperous theoretical results and powerful methods, like the Frank-Wolfe algorithm [30] and the proximal gradient method [32]. In this paper, we use a projected gradient algorithm to solve (P2(\(\theta\))).

**IV. LOW COMPLEXITY SOLUTION**

In the following, we shall introduce Algorithm 1 to solve (P2(\(\theta\))), which is shown below. Step 1 is to solve an \(l_2\) norm estimation of (P0) as an initial point. Step 2 is the projected gradient loop, which is used to find the solution of the smooth optimization (P2(\(\theta\))) for a decreasing sequence of \(\theta\). Finally, based on the obtained backhaul cooperation solution, Step 3 is to find the minimum power consumption solution. An illustrative block diagram of Algorithm 1 is summarized in Figure 3.

\[
\begin{align*}
\text{Step 1} & \quad \text{Initialization (}\ l_2 \text{norm estimation) } & \text{Step 2} & \quad \text{Projected Gradient Loop to solve (P2(\(\theta\))) } & \text{Step 3} & \quad \text{Find the min. power sol. (} W_f^* \text{)}
\end{align*}
\]

![Fig. 3. Block diagram for Algorithm 1.](image)

The \(l_2\) norm estimation of (P0) in Step 1 is actually the problem (OBD), which serves as a rough estimate for the initial point of Algorithm 1. The projected gradient loop in Step 2 is equivalent to repeatedly solve the projection problem (AP), which is shown as follows:

\[
(AP) \quad \min_{\{w_k, w_j\}_{k=1}^K} \sum_{k=1}^K \|w_k - w_k^*\|_2^2
\]

s.t. Constraints in (25),

where \(\{w_k\}_{k=1}^K\) coming from the gradient computation step as illustrated in Figure 3 are the vectors to be projected. To obtain the minimum power solution, an (OBP-AC) is solved in Step 3. Note that (OBP) and (OBP-AC) are not convex due to the SINR constraints. On the other hand, we observe that the projection problem (AP) can be transformed into a pure SDP form, which is more amenable for rank characterization and analysis. We shall first perform the subproblem transformations next.

**A. Transformations of Subproblems in Algorithm 1**

1) Semidefinite Relaxation in (OBD) and (OBP-AC): By using the relation in (21) and (23) to drop the rank constraint, we can relax the non-convex (OBD) into the following form:
where all entries of $O$ are inhomogeneous quadratic, which can be
Similarly, $(OBP-SDP)$ can be relaxed into the following form:
min$_{(W_k)_{k=1}^K} \sum_{k=1}^K \text{Tr}(W_k)$  
\text{s.t.} \quad \text{Tr}(H_k W_k) - \gamma_k \sum_{m \neq k} \text{Tr}(H_k W_m) - \gamma_k \sigma_k^2 \geq 0
\quad W_k \succeq 0, \quad \forall k

Similarly, $(OBP-AC)$ can be relaxed into the following form:
min$_{(W_k)_{k=1}^K} \sum_{k=1}^K \text{Tr}(W_k)$  
\text{s.t.} \quad \text{Constraints in (27).}
\quad \text{Tr}(M_k W_k) = 0, \quad \forall k

2) Transformation of $(AP)$ into SDP Form: Note that the
objective of $(AP)$ is inhomogeneous quadratic, which can be
rewritten as
\[ ||w_k - \tilde{w}_k||_2^2 = \left[ w_k^T \begin{bmatrix} \tilde{w}_k^T \tilde{w}_k & -\tilde{w}_k^T \\ -\tilde{w}_k & I \end{bmatrix} \tilde{w}_k \right] 1 \begin{bmatrix} 1 \\ w_k^T \end{bmatrix} . \]

Let
\[ I_{00} = \begin{bmatrix} 1 & 0^H \\ 0 & 0 \end{bmatrix} , \]
where all entries of $O \in S^N$ and $0 \in C^N$ are 0. Thus $I_{00} \in S^{N+1}$. Let $I \in S^N$,
\[ A_k = \begin{bmatrix} \tilde{w}_k^T \tilde{w}_k & -\tilde{w}_k^T \\ -\tilde{w}_k & I \end{bmatrix} \text{ and } \tilde{H}_k = \begin{bmatrix} 0 & 0^H \\ 0 & H_k \end{bmatrix} . \]

Thus, $(OBP-AC)$ can be transformed into SDP form:
\[ \min_{(W_k)_{k=1}^K} \sum_{k=1}^K \text{Tr}(A_k W_k) \]
\text{s.t.} \quad \text{Constraints in (27).}
\quad \text{Tr}(I_{00} W_k) = 1, \quad \forall k
\quad W_k \succeq 0, \quad \forall k

Note that $(AP)$ and $(AP-SDP)$ are equivalent before dropping
the rank constraint. As such, $(AP)$ is replaced by solving
$(AP-SDP)$ in Algorithm 1. In the following subsection $B$, we
will show that the rank relaxation for $(AP-SDP)$ is always tight
with probability 1.

The overall solution structure to solve $(P0)$ is summarized in
Figure 4.

B. Discussion of the Optimality of Algorithm 1

$(OBP-SDP)$ is a power minimization problem subject to
SINR requirements that is known to have rank 1 optimal solu-
tions by using uplink-downlink duality and Perron-Frobenius
theory for matrices with nonnegative entries \cite{6}. For complete-
tness, we provide a simple proof using the KKT conditions.

Theorem 2 (Strong duality of $(OBP-SDP)$): Suppose the optimal
solution for $(OBP-SDP)$ is $(W_k)_{k=1}^K$, then rank$(W_k) = 1, \forall k = 1, ..., K$.

Proof: The lagrangian dual problem (OBP-SDP) is shown below:

Adding this small perturbation is to ensure that the output of the SDR
solution is rank 1 with probability 1, which will be shown in the proof of
Theorem 3.

Strong duality means that all the optimal solutions are rank 1, i.e.,
the SDR relaxation is tight.

Algorithm 1 Proposed Algorithm

- **Step 1 Initialization:**
  1) Let $(w_k)_{k=1}^K$ be the $l_2$ norm solution, i.e., the solution of the optimization problem (OBP-SDP).
  2) Set the $\theta$ decreasing factor $0 < \eta < 1, \tau > 0$ and the perturbation factor $\phi$.
- **Step 2 Projection gradient loop:** For $j = 1, ..., J$
  1) Let $\theta = \theta^j$.
  2) Projection gradient step: Maximize the objective $F = (F_0(w_1, ..., w_K) - \epsilon \sum_{k=1}^K ||w_k||_2^2)$ on the feasible set using
the steepest ascent method followed by projection onto the feasible set.
    - Find the gradient $\delta_k = \nabla w_k F \in \mathbb{C}^{L_N}$.
    - Let the perturbation be $e(w_k^{j-1}, \mu) = \mu \zeta \omega_j - 1$, where $\zeta \sim \mathcal{U}[-\theta, \theta]$.
    - Let $\tilde{w}_k = w_k^{j-1} + \mu \delta_k - e(w_k^{j-1}, \mu)$, $\forall k$.
    - Project $\tilde{w}_k$ back onto the feasible set to get $w_k^j$, i.e., equivalent to solving the projection problem (AP) (26).
  3) $\theta$ updating step: set $||\sum_{k=1}^K \delta_k(w_k^j - w_k^{j-1})|| \geq \tau \theta^{j-1}$, then set $\theta^j = \theta^{j-1}$; otherwise, choose $\theta^j = \eta \theta^{j-1}$.
- **Step 3 Find the minimum power solution:** Obtain the positions of the zero entries in $w_k^j$ to form $m_k, \forall k$.
  1) Use $(m_k)_{k=1}^K$ to solve an (OBP-AC-SDP) and obtain the solution $(\tilde{w}^j_k)_{k=1}^K$.
  2) **The final solution is given by $(\tilde{w}^j_k)_{k=1}^K$.**

\[ \text{(OBP-SDP)} \]
\[ \min_{(W_k)_{k=1}^K} \sum_{k=1}^K \text{Tr}(W_k) \quad \text{(27)} \]
\text{s.t.} \quad \text{Tr}(H_k W_k) - \gamma_k \sum_{m \neq k} \text{Tr}(H_k W_m) - \gamma_k \sigma_k^2 \geq 0
\quad W_k \succeq 0, \quad \forall k

\[ \text{(OBP-AC-SDP)} \]
\[ \min_{(W_k)_{k=1}^K} \sum_{k=1}^K \text{Tr}(W_k) \quad \text{(28)} \]
\text{s.t.} \quad \text{Constraints in (27).}
\quad \text{Tr}(M_k W_k) = 0, \quad \forall k

\[ \text{(AP-SDP)} \]
\[ \min_{(W_k)_{k=1}^K} \sum_{k=1}^K \text{Tr}(A_k W_k) \quad \text{(32)} \]
\text{s.t.} \quad \text{Tr}(H_k W_k) - \gamma_k \sum_{m \neq k} \text{Tr}(H_k W_m) - \gamma_k \sigma_k^2 \geq 0
\quad W_k \succeq 0, \quad \forall k

Strong duality with probability 1.

**Theorem 4 (Strong duality of (OBP-AC-SDP))**: Suppose the optimal solution for (OBP-AC-SDP) is \( \{W_k^*\}_{k=1}^K \), then rank\( (W_k^*) = 1, \forall k = 1, ..., K \) with probability 1.

**Proof**: See Appendix B.

**Remark**: In the literature, not all SDR problems have rank 1 characterizations and even if some problems are shown to have rank 1 solution in the numerical sense, it is still highly challenging to find an analytical support [33]. In general, the rank 1 justification for SDR problems are case specific and the rank 1 property from one problem cannot be directly generalized to another one. We summarize the difference between our work and the existing works [6], [20] and [35] in Table 1.

**Theorem 5 (Convergence of Algorithm 1 to the stationary point of (P0))**: The limiting output \( \{\hat{w}_k\}_{k=1}^K \) of algorithm 1 is a stationary point of the non-smooth problem (P0) with probability 1. Furthermore, all the intermediate outputs are rank 1 almost surely.

**Proof**: See Appendix C.

### C. Implementation Considerations

In this subsection, we shall present the implementation considerations and complexity analysis for Algorithm 1 as follows.

**Implementation Considerations**: Note that the proposed algorithm is implemented in a centralized way. In practical networks, all the BSs can be connected to a central controller or one of BSs can act as the central controller through the backhaul. The CP collects all the required information such as CSI \( h_k \) and SINR \( \gamma_k \) and then computes the beamforming parameters in a centralized manner. Hence, the beamforming solutions could be distributed to each corresponding BSs through the backhaul.

**Complexity Analysis**: Note that (OBP-SDP), (AP-SDP) and (OBP-AC-SDP) could be solved by SeDuMi by means of an interior point method and they have the same worst-case complexity order \( O((K^{3.5}L^{6.5}N^{6.5})\log(1/\varepsilon)) \), where \( \varepsilon \) represents the accuracy of the solution at the algorithm’s termination. Since (OBP-SDP) and (OBP-AC-SDP) are solved only once for initialization and finalization, respectively, the complexity of the proposed algorithm mainly comes from the repeated executions of (AP-SDP). Hence, the worst-case complexity of the proposed algorithm is \( O((K^{3.5}L^{6.5}N^{6.5})\log(1/\varepsilon)\log(1/\theta_{\min})) \), where \( \theta_{\min} \) represents the accuracy of the approximation to the non-smooth \( l_0 \) norm. This complexity is substantially lower than the brute force search complexity \( O(2^{KN}) \).

### V. NUMERICAL RESULTS

**A. Setting of the Algorithm Parameters**

The initial value \( \theta^1 \) is set to two times the maximum absolute value of the entries in the vectors \( \{w_k^0\}_{k=1}^K \). The existence of rank 1 optimal solution means that there exists a rank 1 optimal solution among all the optimal solutions, which may have various rank profiles. It is considered as a weaker result than strong duality.
reason is that this value of θ is virtually like infinity for all the entries in \( \{w_k^{(0)}\}_{k=1}^K \). The update of θ is controlled by τ and η. Note that we do not need to wait for the convergence of the internal loop. For a smaller value of τ, more iterations in the internal loop are needed. Depending on the network configuration, the choice of τ should be empirically between \( KLN \sqrt{\gamma}/3 \) and \( KLN \sqrt{\gamma} \), where γ is SINR. The θ decreasing factor η is typically chosen from 0.7 ∼ 0.95 and it is set to 0.9 in the simulation. A larger η requires more iterations and it is more likely to get the global optimal. The stepsize \( \Delta \) is chosen to be decreasing with the smooth parameter θ, which is because the smooth function fluctuates more with smaller θ; therefore smaller \( \Delta \) should be applied and it is set to \( 2(\theta)^2 \). The small perturbation factor \( \varphi \) is chosen to be 0.0001 to have neglecting effect on the gradient. The algorithm will stop once θ is below a sufficiently small threshold value \( \theta_{min} \).

The summary of the algorithm parameters is given in Table II.

### Table II

**Summary of Algorithm Parameters**

| Para. | Function | Typical values |
|-------|----------|----------------|
| \( \theta^1 \) | initial value of \( \theta \) | \( (2 \sim 4) \times \max_{n,k,l} |w_{n,k,l}^{(0)}| \) |
| \( \eta \) | \( \theta \) decreasing factor | 0.7 ∼ 0.95 |
| \( \theta_{min} \) | threshold value of \( \theta \) | 0.0001 ∼ 0.001 |
| \( \tau \) | \( \theta \) update control | \( KLN \sqrt{\gamma}/3 \sim KLN \sqrt{\gamma} \) |
| \( \mu^\gamma \) | stepsize | \( 2(\theta)^2 \sim 3(\theta)^2 \) |
| \( \varphi \) | perturbation factor | 0.0001 ∼ 0.001 |

### C. Random Network Model and Baseline Setups

**Random Network Model**

Consider a Poisson Point Process (PPP) model for the BSs and MSs in a 1km*1km square area with the BS density given by \( \lambda_{BS} = 4/\text{km}^2 \) and MS density given by \( \lambda_{MS} = 8/\text{km}^2 \). Consider that each BS has \( L = 2 \) antennas and each MS has single antenna. The channel coefficient between the \( l \)-th antenna of the \( n \)-th BS and the \( k \)-th MS is:

\[
h_{n,k,l} = \Gamma_{n,k,l} \sqrt{G} d_{n,k}^{-\frac{\xi}{2}},
\]

where \( d_{n,k} \) is the distance between the \( n \)-th BS and the \( k \)-th MS. \( G \) is the BS antenna power gain, which is assumed to be 9dB. \( \xi \) is the path-loss exponent, and \( \beta \) is the path-loss constant. \( \Gamma_{n,k,l} \sim \mathcal{N}(0, 1) \) is the complex Gaussian coefficient, which models the large scale fading (shadowing). \( \Gamma_{n,k,l}^t \sim \mathcal{NC}(0, 1) \) is the complex Gaussian coefficient, which models the small scale fading. For the pathloss model, the 3GPP Long Term Evolution (LTE) standard has been used:

\[
PL_{n,k}^{dB} = 148.1 + 37.6 \log_{10}(d_{n,k}^{km}).
\]

For each random network configuration, \( 10^3 \) runs are used to simulate the PPP model location. For each PPP model location, \( 10^3 \) channel realizations are used to simulate the fading channel.

For different network configurations, \( \epsilon \) should be properly chosen to yield the desired tradeoff. Generally, a smaller \( \epsilon \) promotes a more sparse solution, but with potentially high power consumption, while a larger \( \epsilon \) takes power consumption as a priority, which may result in more cooperations. It is worth noting that when \( \epsilon \) equals 0, the design problem only focuses on the minimization of the active backhaul links, regardless of the transmit power consumption. For a very large \( \epsilon \), the weight of the power consumption (\( l_2 \) norm part) overwhelms the \( l_0 \) norm counterpart; thus the problem reduces to an (OBP) problem.

**B. Impact of parameter \( \epsilon \) on the tradeoff between Cooperation and Power Consumption**

For different network configurations, \( \epsilon \) should be properly chosen to yield the desired tradeoff. Generally, a smaller \( \epsilon \) promotes a more sparse solution, but with potentially high power consumption, while a larger \( \epsilon \) takes power consumption as a priority, which may result in more cooperations. It is worth noting that when \( \epsilon \) equals 0, the design problem only focuses on the minimization of the active backhaul links, regardless of the transmit power consumption. For a very large \( \epsilon \), the weight of the power consumption (\( l_2 \) norm part) overwhelms the \( l_0 \) norm counterpart; thus the problem reduces to an (OBP) problem.

**Random Network Model**

Consider a Poisson Point Process (PPP) model for the BSs and MSs in a 1km*1km square area with the BS density given by \( \lambda_{BS} = 4/\text{km}^2 \) and MS density given by \( \lambda_{MS} = 8/\text{km}^2 \). Consider that each BS has \( L = 2 \) antennas and each MS has single antenna. The channel coefficient between the \( l \)-th antenna of the \( n \)-th BS and the \( k \)-th MS is:

\[
h_{n,k,l} = \Gamma_{n,k,l} \sqrt{G} d_{n,k}^{-\frac{\xi}{2}},
\]

where \( d_{n,k} \) is the distance between the \( n \)-th BS and the \( k \)-th MS. \( G \) is the BS antenna power gain, which is assumed to be 9dB. \( \xi \) is the path-loss exponent, and \( \beta \) is the path-loss constant. \( \Gamma_{n,k,l} \sim \mathcal{N}(0, 1) \) is the complex Gaussian coefficient, which models the large scale fading (shadowing). \( \Gamma_{n,k,l}^t \sim \mathcal{NC}(0, 1) \) is the complex Gaussian coefficient, which models the small scale fading. For the pathloss model, the 3GPP Long Term Evolution (LTE) standard has been used:

\[
PL_{n,k}^{dB} = 148.1 + 37.6 \log_{10}(d_{n,k}^{km}).
\]

For each random network configuration, \( 10^3 \) runs are used to simulate the PPP model location. For each PPP model location, \( 10^3 \) channel realizations are used to simulate the fading channel.

For different network configurations, \( \epsilon \) should be properly chosen to yield the desired tradeoff. Generally, a smaller \( \epsilon \) promotes a more sparse solution, but with potentially high power consumption, while a larger \( \epsilon \) takes power consumption as a priority, which may result in more cooperations. It is worth noting that when \( \epsilon \) equals 0, the design problem only focuses on the minimization of the active backhaul links, regardless of the transmit power consumption. For a very large \( \epsilon \), the weight of the power consumption (\( l_2 \) norm part) overwhelms the \( l_0 \) norm counterpart; thus the problem reduces to an (OBP) problem.

**Random Network Model**

Consider a Poisson Point Process (PPP) model for the BSs and MSs in a 1km*1km square area with the BS density given by \( \lambda_{BS} = 4/\text{km}^2 \) and MS density given by \( \lambda_{MS} = 8/\text{km}^2 \). Consider that each BS has \( L = 2 \) antennas and each MS has single antenna. The channel coefficient between the \( l \)-th antenna of the \( n \)-th BS and the \( k \)-th MS is:

\[
h_{n,k,l} = \Gamma_{n,k,l} \sqrt{G} d_{n,k}^{-\frac{\xi}{2}},
\]

where \( d_{n,k} \) is the distance between the \( n \)-th BS and the \( k \)-th MS. \( G \) is the BS antenna power gain, which is assumed to be 9dB. \( \xi \) is the path-loss exponent, and \( \beta \) is the path-loss constant. \( \Gamma_{n,k,l} \sim \mathcal{N}(0, 1) \) is the complex Gaussian coefficient, which models the large scale fading (shadowing). \( \Gamma_{n,k,l}^t \sim \mathcal{NC}(0, 1) \) is the complex Gaussian coefficient, which models the small scale fading. For the pathloss model, the 3GPP Long Term Evolution (LTE) standard has been used:

\[
PL_{n,k}^{dB} = 148.1 + 37.6 \log_{10}(d_{n,k}^{km}).
\]

For each random network configuration, \( 10^3 \) runs are used to simulate the PPP model location. For each PPP model location, \( 10^3 \) channel realizations are used to simulate the fading channel.

Five baseline schemes are considered for comparison

1. Baseline 1 is a directional cooperation scheme which greedily chooses the links by evaluating the strength of the channel response \( [36] \). A deflation heuristic is used to minimize the selected links.

2. Baseline 2 is an opportunistic cooperation scheme with suitable switching between the full cooperation model and the coordination mode according to the users’ SINR constraints \( [1] \).

3. Baseline 3 is a dynamic clustering scheme that dynamically forms user-centric clusters for each MS to mitigate interference \( [12] \). A greedy algorithm is used in this scheme by starting from the lowest number of cooperating clusters to the full cooperation scenario.

4. Baseline 4 is an iteratively link removal algorithm in \( [21] \) that iteratively removes the links that correspond to the smallest link transmit power based on the \( l_2 \) norm relaxation.

5. Baseline 5 is a reweighted \( l_1 \) norm minimization algorithm in \( [21] \) that is based on \( l_1 \) norm relaxation.
The algorithm solves a series of $l_1$-norm minimization problems to obtain the minimum backhaul cooperation solution.

D. Performance comparison with Baselines

Figure 5 investigates the impact of $\epsilon$, which controls the tradeoff between cooperation and power consumption in the proposed algorithm (Note the curves are not plotted together due to the different operating regions of the respective schemes). From Figure 5, it is readily observed that the curve of the proposed algorithm is strictly lower than the other baselines. In other words, when fixing the number of cooperating BSs per MS, the proposed algorithm consumes less transmit power than the baselines. When fixing the transmit power, the proposed algorithm requires less number of cooperations than the baselines.

Figure 6 and Table III illustrate the backhaul cost and power consumption respectively for different schemes versus the SINR requirements $\gamma_k$ (assume $\gamma_k$ are equal for the users). Three values of $\epsilon$ are considered, $\epsilon = 0$, $\epsilon = 0.1$ and $\epsilon = 0.5$. Specifically, when $\epsilon = 0$, the proposed algorithm achieves the minimum backhaul cost at the expense of more power consumption compared with $\epsilon = 0.1$ and $\epsilon = 0.5$. From Figure 6, we observe that the backhaul cost increases as the SINR increases. Moreover, we see that the proposed scheme always have a better performance than the baselines both in terms of the backhaul cost and the power consumption.

Figure 7 and Table IV investigate the performance comparison between the proposed algorithm and the baselines as MS density varies in terms of backhaul cost and power consumption, respectively. Similar to the observations we have in Figure 6 and Table III, the proposed algorithm always has substantial gains over the various baselines in respect to both backhaul cost and power consumption. Moreover, we observe that the backhaul cost increases as the MS density increases.

By jointly processing the signals at the CP, the proposed scheme is able to choose the cooperation links on a larger scale and achieve a better balanced solution. It is worth noting that the proposed algorithm outperforms the Cluster-MIMO scheme, which means the inter-cluster interference is carefully taken into consideration when jointly designing the asymmetric cooperation and beamforming. It is also noteworthy that the proposed algorithm achieves substantial gains over the reweighted $l_1$ norm minimization and iterative link removal algorithms in [21]. The intuitions behind are as follows. Although $l_1$ norm is the best known convex approximation to $l_0$ norm, the smoothed $l_0$ norm used in this paper is the closest approximation since it is exactly the same as $l_0$ norm when the smooth parameter tends to 0. This better approximation and formulation may lead to a better solution over the $l_1$ norm based algorithms. On the other hand, the iterative link removal approach proposed is only a heuristic algorithm that gradually removes the links purely based on the transmit power.

E. Performance comparison with Full search

Due to the intolerable full search complexity for large problem size, we consider a small random network consists

![Graph](image-url)
of 3BSs and 3MSs. The SINR threshold $\gamma_k$ is set to 20dB for all the users. Table III compares the performance of the proposed algorithm with combinatorial full search in terms of average number of cooperating BSs per MS (denoted as Avg. Coop) and average transmit power (denoted as Avg. Pwr). Three different values of $\epsilon$ are considered, i.e., 0, 0.1, 0.5.

Overall the proposed algorithm performs reasonably close to the optimal case given the tremendous saving in respect to complexity. In this simple setup, full search needs to solve 465 SDP iterations while the proposed scheme only requires 35.2 on average.

### VI. CONCLUSION

In this work, we propose a novel scheme to reduce the backhaul loading as well as the power consumption in MIMO cellular networks. We model the backhaul cost as the number of direct cooperation links and formulate a combinatorial optimization problem that minimizes the $l_0/l_2$ norm of the beamforming vectors subject to SINR constraints at all the users. We approximate the non-smooth combinatorial part with a smooth function and relax the non-convex SINR constraints using SDR. A practical and efficient algorithm is introduced to solve the approximated problem. The proposed algorithm guarantees a rank 1 solution with probability 1 and convergence to the stationary point. The effectiveness of the proposed algorithm is demonstrated via extensive simulations.

### APPENDIX A

#### PROOF OF THEOREM 3

Before proving Theorem 3, we introduce the following lemmas.

**Lemma 1**: (Proposition 4, [8]) If $A \succeq 0, B \succeq 0$ and $c$ is in the range of $A$, then

$$c^H A c \geq c^H (A + B) c,$$

with equality if and only if $B (A + B) c = 0$.

**Lemma 2**: If $A \succeq 0$ and $A (I + A)^{-1} c = 0$, then

$$Ac = 0 \quad \text{and} \quad (I + A)^{-1} c = c$$

**Proof**: Since $(I + A)$ is a full rank matrix, there always exists a $y$ such that $(I + A)y = c, \forall c$. Therefore, $A(I + A)^{-1}(I + A)y = 0$, i.e. $Ay = 0$. This means $y \in \mathcal{N}(A)$ and $c = y = (I + A)^{-1} c$. □

**Proof of Theorem 3**: We first relax the constraint $\text{Tr}(I_{00}\tilde{W}_k) = 1$ in (AP-SDP) to $\text{Tr}(I_{00}\tilde{W}_k) \leq 1$, resulting in the relaxed problem (AP-SDP-Relaxed):

$$\begin{align*}
\min_{\{\tilde{W}_k\}_{k=1}^K} & \quad \sum_{k=1}^K \text{Tr}(A_k \tilde{W}_k) \\
\text{s.t.} & \quad \text{Tr}(H_k \tilde{W}_k) - \gamma_k \sum_{m \neq k} \text{Tr}(H_k \tilde{W}_m) - \gamma_k \sigma_k^2 \geq 0 \\
& \quad \text{Tr}(I_{00} \tilde{W}_k) \leq 1, \quad \forall k \\
& \quad \tilde{W}_k \succeq 0, \quad \forall k
\end{align*}$$

We shall first prove that the SDR of (AP-SDP-Relaxed) is always tight given the sufficient condition that $H_k \tilde{w}_k \neq 0, \forall k$. Let the optimal value of (AP-SDP) be $p^*$ and the optimal value of (AP-SDP-Relaxed) be $d^*$. Then $p^* \geq d^*$. The lagrangian dual problem of (AP-SDP-Relaxed) is shown below. Let $\lambda_k, \varphi_k$ and $Z_k$ be the lagrangian multipliers for the SINR inequality, the trace inequality and the PSD constraint, respectively.

$$\begin{align*}
\max_{\{\lambda_k, \varphi_k\}_{k=1}^K} & \quad \sum_{k=1}^K (\lambda_k \gamma_k \sigma_k^2 - \varphi_k) \\
\text{s.t.} & \quad Z_k = A_k - \lambda_k H_k + \sum_{m \neq k} \lambda_m \gamma_m H_m + \varphi_k I_{00} \succeq 0 \\
& \quad \lambda_k \geq 0, \quad \forall k \\
& \quad \varphi_k \geq 0, \quad \forall k
\end{align*}$$

$Z_k$ can be written in the following matrix form:

$$Z_k = \begin{bmatrix}
\tilde{w}_k^H \tilde{w}_k + \varphi_k \\
-\tilde{w}_k^H \tilde{I} - \sum_{m \neq k} \lambda_m \gamma_m H_m - \lambda_k H_k
\end{bmatrix} \succeq 0.$$
If \( \varphi_k > 0 \), then by Schur complement

\[
A_k + \varphi_k I_{m0} = \begin{bmatrix}
\tilde{w}_k^H w_k + \varphi_k & -w_k^H \\
-w_k & I
\end{bmatrix} > 0.
\]

Since \( \tilde{w}_k^H w_k + \varphi_k - w_k^H w_k > 0 \). Thus \( A_k + \varphi_k I_{m0} + \sum_{m \neq k} \lambda_m \gamma_m H_m \) is also a full rank matrix. By KKT conditions, we know,

\[
A_k + \sum_{m \neq k} \lambda_m \gamma_m \tilde{H}_m + \varphi_k I_{m0} - \lambda_k \tilde{H}_k - Z_k = 0,
\]

(43)

\[
\tilde{W}_k Z_k = 0.
\]

(44)

Premultiplying the two sides of (43) by \( \tilde{W}_k \), and making use of (44), we get

\[
\tilde{W}_k (A_k + \sum_{m \neq k} \lambda_m \gamma_m \tilde{H}_m + \varphi_k I_{m0}) = \lambda_k \tilde{W}_k H_k.
\]

(45)

Now the following relation holds:

\[
\text{rank}(\tilde{W}_k) = \text{rank}(\tilde{W}_k (A_k + \sum_{m \neq k} \lambda_m \gamma_m \tilde{H}_m + \varphi_k I_{m0}))
\]

\[
= \text{rank}(\lambda_k \tilde{W}_k H_k)
\]

\[
\leq \min(\text{rank}(H_k), \text{rank}(\tilde{W}_k)),
\]

(46)

(47)

(48)

where (46) is due to \( A_k + \sum_{m \neq k} \lambda_m \gamma_m \tilde{H}_m + \varphi_k I_{m0} \) being a full rank matrix. and (48) follows from a basic rank inequality [37]. Because \( H_k \) is rank 1, we obtain \( \text{rank}(\tilde{W}_k) \leq 1 \). Since \( \tilde{W}_k \neq 0 \), the rank of \( W_k \) must be 1. For \( \varphi_k = 0 \), because \( Z_k \geq 0 \), by Generalized Schur Complement (for a singular bottom right corner matrix), we have

\[
\tilde{w}_k^H w_k + \varphi_k - w_k^H (I + \sum_{m \neq k} \lambda_m \gamma_m H_m - \lambda_k H_k)^{-1} \tilde{w}_k \geq 0,
\]

(49)

\[
\tilde{w}_k \in \mathcal{R}(I + \sum_{m \neq k} \lambda_m \gamma_m H_m - \lambda_k H_k)
\]

(50)

\[
I + \sum_{m \neq k} \lambda_m \gamma_m H_m - \lambda_k H_k \succeq 0.
\]

(51)

Therefore, by Lemma 1,

\[
\tilde{w}_k^H w_k \geq \tilde{w}_k^H (I + \sum_{m \neq k} \lambda_m \gamma_m H_m - \lambda_k H_k)^{-1} \tilde{w}_k
\]

(52)

\[
\geq \tilde{w}_k^H (I + \sum_{m \neq k} \lambda_m \gamma_m H_m)^{-1} \tilde{w}_k. \tag{1}
\]

\[
\tilde{w}_k^H w_k \geq \tilde{w}_k^H (I + \sum_{m \neq k} \lambda_m \gamma_m H_m)^{-1} \tilde{w}_k. \tag{2}
\]

The conditions for the equality to hold in (1) and (2) are

\[
\lambda_k H_k (I + \sum_{m \neq k} \lambda_m \gamma_m H_m)^{-1} \tilde{w}_k = 0
\]

(53)

\[
\lambda_k H_k (I + \sum_{m \neq k} \lambda_m \gamma_m H_m)^{-1} \tilde{w}_k = 0
\]

(54)

Note that if (54) holds, then by Lemma 2, \((I + \sum_{m \neq k} \lambda_m \gamma_m H_m)^{-1} \tilde{w}_k = \tilde{w}_k\) and \((I + \sum_{m \neq k} \lambda_m \gamma_m H_m) \tilde{w}_k = 0\). Thus, (53) can be written as \(\lambda_k H_k \tilde{w}_k = 0\). If the equality in (2) holds (i.e., (54)), then (53) has to hold. Given the sufficient condition that \(H_k \tilde{w}_k \neq 0\), \(\lambda_k\) has to be 0. Then \(Z_k = A_k + \sum_{m \neq k} \lambda_m \gamma_m H_m\) and \(\text{rank}(Z_k) = \nu\); thus \(\text{rank}(W_k) = 1\).

If (54) does not hold, then \(A_k + \sum_{m \neq k} \lambda_m \gamma_m \tilde{H}_m\) is a full rank matrix. Using the same relation in (46), we have \(\text{rank}(W_k) = 1\).

Summing the statements above, the solution of (AP-SDP-Relaxed) must be rank 1. As a matter of fact, the sufficient condition \(H_k \tilde{w}_k \neq 0\) holds with probability 1 in Algorithm 1. Since \(\tilde{w}_k = w_k + \mu^j \delta_k - \mu^j \gamma \tilde{w}_k^{-1}\), if \(H_k \tilde{w}_k = 0\) then \(\mu^j H_k \tilde{w}_k^{-1} = H_k (w_k^{-1} + \mu^j \delta_k)\). However, the right hand side is predetermined and \(H_k \tilde{w}_k^{-1} \neq 0\), the above equality holds with probability 0 when \(\zeta\) is randomly chosen. When \(\mu^j \to 0\), the right hand side reduces to \(H_k \tilde{w}_k^{-1}\), which is nonzero as well. Therefore, the sufficient condition holds with probability 1 in Algorithm 1.

Also note that \(W_k \geq 0\) and \(\text{Tr}(I_{m0} W_k) \leq 1\), then \(W_k\) should behave in the following structure:

\[
\tilde{W}_k = \begin{bmatrix}
z^2 & zw_k^H \\
z w_k & w_k H_k
\end{bmatrix}
\]

(55)

with \(0 < z \leq 1\). Thus, from the KKT conditions \(Z_k \tilde{W}_k = 0\) and \(\varphi_k (\text{Tr}(I_{m0} \tilde{W}_k) - 1) = 0\), we have

\[
(I + \sum_{m \neq k} \lambda_m \gamma_m H_m - \lambda_k H_k) w_k = z \bar{w}_k,
\]

(56)

\[
z^2 (\tilde{w}_k^H \tilde{w}_k + \varphi_k) = z \tilde{w}_k^H w_k
\]

(57)

\[
\varphi_k (z - 1) = 0.
\]

(58)

Thus if \(\varphi_k > 0\), by (58), \(z = 1\). If \(\varphi_k = 0\), from (57) we know \(z^2 \tilde{w}_k^H \tilde{w}_k = z \tilde{w}_k^H w_k\). Note \(\tilde{w}_k^H w_k\) is a positive scalar, which can be easily observed by premultiplying \(w_k\) to (56).

\(\tilde{W}_k\) is a feasible solution to (AP-SDP-Relaxed) \(\forall 0 < z \leq 1\), since the choice of \(z\) will not affect the SINR constraint and positive semidefinite requirement. By generalized Schur complement, \(z^2 \geq (zw_k^H) (w_k H_k) (zw_k^H)^{-1}\), which can be simplified as \(1 \geq w_k^H (w_k H_k)^{-1} w_k\). The equality holds and obviously \(\forall 0 < z \leq 1\).

In (AP-SDP-Relaxed), the objective can be written as

\[
\text{Tr}(A_k \tilde{W}_k) = z^2 \tilde{w}_k^H \tilde{w}_k - z \tilde{w}_k^H w_k - z \tilde{w}_k^H w_k + z w_k H_k w_k
\]

(59)

Since \(z^2 \tilde{w}_k^H \tilde{w}_k = z \tilde{w}_k^H w_k = z \tilde{w}_k^H w_k\), then \(\text{Tr}(A_k \tilde{W}_k) = -z \tilde{w}_k^H \tilde{w}_k + w_k H_k w_k = -z \tilde{w}_k^H \tilde{w}_k + w_k H_k w_k \geq -\tilde{w}_k^H \tilde{w}_k + w_k H_k w_k\), with the equality holding when \(z = 1\). Therefore, \(z\) must be 1 for the optimal solution. Thus, \(\tilde{W}_k\) is also the optimal solution for (AP-SDP) and \(p^* = d^*\), which implies that the problems (AP-SDP) and (AP-SDP-Relaxed) are equivalent. Therefore, the optimal solution of (AP-SDP) must be rank 1. □

**Appendix B**

**Proof of Theorem 4**

Proof: The lagrangian dual problem of (OBP-AC-SDP) is shown below. \(\lambda_k, \varphi_k\), and \(Z_k\) are the lagrangian multipliers for the SINR inequality, the trace equality and the PSD constraint,
respectively. 

\[
\begin{align*}
\text{(OBP-AC-SDP-Dual)} & \\
\max_{\{\lambda_k, \varphi_k\}_{k=1}^K} & \quad \sum_{k=1}^K \lambda_k \gamma_k \theta_k^2 \\
\text{s.t.} & \quad Z_k = I - \lambda_k H_k + \sum_{m \neq k} \lambda_m \gamma_m H_m + \varphi_k M_k \geq 0 \\
& \quad \lambda_k \geq 0, \quad \forall k
\end{align*}
\]

(60)

Note that the problem (OBP-AC-SDP-Dual) and (OBP-SDP-Dual) (which is the OBP with full cooperation) have the same objective function and only differs in the term \(\varphi_k M_k\) in the constraints. Suppose \(\{\lambda_k^*\}_{k=1}^K\) and \(\{\gamma_k\}_{k=1}^K\) are the optimal solutions of (OBP-AC-SDP-Dual) and \(\{\lambda_k^*\}_{k=1}^K\) and \(\{\gamma_k\}_{k=1}^K\) are the optimal solutions of (OBP-SDP-Dual) respectively. \(\{\lambda_k^*\}_{k=1}^K\) and \(\{\lambda_k^*\}_{k=1}^K\) with \(\varphi_k^* M_k^*\) will be smaller or equal to the objective value of \(\{\lambda_k^*\}_{k=1}^K\).

Notice that \(M_k\) is a predetermined diagonal matrix with diagonal elements being \(1\) or \(0\). Let the null space vector of \(I - \lambda_k^0 H_k + \sum_{m \neq k} \lambda_m^* \gamma_m H_m\) be \(y_k\). Thus, \(y_k\) depends on the i.i.d. channel realizations \(H_k\). From this fact, \(y_k\) is continuously distributed in the \(N\)-dimensional space and consequently, the probability that \(y_k\) falls in the null space of \(M_k\) is zero, i.e., \(\text{Pr}(\text{null space of } M_k) = 0\). The case of \(M_k = 0\) is degenerated to (34).

If \(\varphi_k > 0\), under the condition that \(M_k y_k = 0\) with probability 0, then the matrix \(I - \lambda_k^0 H_k + \sum_{m \neq k} \lambda_m^* \gamma_m H_m + \varphi_k M_k \geq 0\) with probability 1. Thus, \(\exists \theta > 0\) such that

\[
I - \lambda_k^0 H_k + \sum_{m \neq k} \lambda_m^* \gamma_m H_m + \varphi_k M_k - \theta H_k \geq 0
\]

(62)

Then \(\lambda_k^* = \lambda_k^0 + \theta > \lambda_k^0\) and this \(\lambda_k^*\) is still a feasible solution of (60), but with a larger objective value than \(\{\lambda_k^*\}_{k=1}^K\).

Therefore, \(\varphi_k > 0\) when achieving the optimal solution. Then \(I + \sum_{m \neq k} \lambda_m \gamma_m H_m + \varphi_k M_k\) is a full rank matrix. Thus \(\text{rank}(I + \sum_{m \neq k} \lambda_m \gamma_m H_m + \varphi_k M_k - \lambda_k H_k) = N - 1\). Then by the KKT condition that \(Z_k W_k = 0\), \(\text{rank}(W_k) \leq 1\). Since \(W_k \neq 0\), \(\text{rank}(W_k) = 1\) with probability 1.

**APPENDIX C**

**PROOF OF THEOREM 5**

**Proof:** Let \(x = [w_1^T \ w_2^T ... \ w_j^T]^T\) and \(\{x^j\}\) be the intermediate outputs of algorithm 1 in the \(j\)-th iteration of the smoothing parameter \(\theta\).

Let \(C\) denote the feasible convex set of (P2(\(\theta\))), shown at the top of the next page.

Since \(g(\cdot, \theta)\) (defined in Section III. A.) is a smooth function, a sequence of feasible points \(\{x^j\}\) in \(C\) will be generated using the projected gradient method:

\[
x^{j+1} = P_C(x^j - \mu^j \nabla g(x^j) + e(x^j, \mu^j))
\]

(64)

for solving

\[
\min_{x \in C} g(x, \theta),
\]

(65)

where \(\mu^j\) is the step size, \(e(x^j, \mu^j)\) is the small perturbation tending 0 when \(\mu^j \to 0\) and \(P_C\) denotes the projection onto the set \(C\).

If the sequence \(\{x^j\}\) satisfies

\[
\lim_{j \to \infty} \|\nabla g(x^j, \theta)^T (x^{j+1} - x^j)\| = 0,
\]

(66)

then \(x^{j+1}\) is a stationary point [30].

Denote \(J = \{j|\theta^{j+1} = \theta^j\}\). If \(J\) is finite, then there exists an integer \(\bar{j}\) such that for all \(j > \bar{j}\),

\[
\|\nabla g(x^j, \theta)^T (x^{j+1} - x^j)\| \geq \tau \theta^j,
\]

and \(\theta^k = \bar{\theta}\) for all \(j \geq \bar{j}\) in the \(\theta\) updating step of the proposed algorithm. (Note \(\|\sum_{k=1}^J \delta_k (w_k^0 - w_k^{j+1})\|\) is equivalent to \(\|\nabla g(x^j, \theta)^T (x^{j+1} - x^j)\|\).) We know from above that when solving the smooth optimization problem over a convex set

\[
\min_{x \in C} g(x, \bar{\theta}),
\]

(68)

the sequence of generated feasible points \(\{x^j\}\) will converge to a stationary point \(^7\) and thus satisfies

\[
\lim_{j \to \infty} \|\nabla g(x^j, \bar{\theta})^T (x^{j+1} - x^j)\| = 0,
\]

(69)

which contradicts (67). This shows that \(J\) must be infinite and \(\lim_{j \to \infty} \theta^j = 0\).

Since \(J\) is infinite, we can assume that \(J = \{j_0, j_1, ...\}\) with \(j_0 < j_1 < ...\). Then we have

\[
\lim_{i \to \infty} \|\nabla g(x^{j_i}, \theta^i)^T (x^{j_i+1} - x^{j_i})\| \leq \tau \lim_{i \to \infty} \theta^i = 0.
\]

(70)

Let \(x^j\) be an accumulation point of \(\{x^{j_i+1}\}\), then \(x^j\) is a stationary point of (P2(\(\theta\))) when \(\theta\) tends to 0.

From Theorem 2 and 3, we know that (OBP-SDP) has strong duality and (AP-SDP) have strong duality with probability 1. Therefore, the overall SDPs in Algorithm 1 have strong duality with probability 1, which implies the SDR of (P1(\(\theta\))) is tight with probability 1. Thus, the stationary point of (P2(\(\theta\))) is the stationary point of (P1(\(\theta\))), which is also the stationary point of the original non-smooth problem (P0) by Definition 1. Since \(l_0\) norm is size-insensitive, an (OBP-AC-SDP) deactivating the zero entries has to be solved to obtain the stationary point with minimum power, \(x^j\). From Theorem 4, we know that (OBP-AC-SDP) has strong duality with probability 1. Thus, this \(x^j\) is the stationary point with minimum required power of the original non-smooth problem (P0).

**REFERENCES**

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu. Multi-cell mimo cooperative networks: A new look at interference. IEEE Journal on Selected Areas in Communications, 28(9):1380–1408, december 2010.

\(^7\)Note that because \(e(w_k^{j+1} - \mu^j)\) → 0 when \(\mu^j \to 0\), this small perturbation will not affect the convergence of the projection gradient algorithm, as shown in [38] and [39].
\[
C = \left\{ \mathbf{x} = [\mathbf{w}_1^T \mathbf{w}_2^T \ldots \mathbf{w}_K^T]^T \mid \text{Tr}(\mathbf{H}_k \mathbf{w}_k^H) - \gamma_k \sum_{m \neq k} \text{Tr}(\mathbf{H}_k \mathbf{w}_m^H) - \gamma_k \sigma_k^2 \geq 0, \quad \text{and} \quad \begin{bmatrix} \mathbf{w}_k^H & \mathbf{w}_k & 1 \end{bmatrix} \geq 0, \forall k = 1, \ldots, K \right\}
\] (63)