Uplink Rate Region of a Coordinated Cellular Network with Distributed Compression

Aitor del Coso and Sebastien Simoens†
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Castelldefels, Spain
†Motolora Labs, Paris, France
email: aitor.delcoso@cttc.es, simoens@motorola.com

***Abstract*— We consider the uplink of a backhaul-constrained coordinated cellular network. That is, a single-frequency network with N multi-antenna base stations (BSs) that cooperate in order to decode the users' data, and that are linked by means of a lossless backhaul with limited capacity. To implement cooperation among receivers, we propose distributed compression: the cooperative BSs, upon receiving their signals, compress them using a distributed Wyner-Ziv code. Then, they send the compressed vectors to the central unit (also a BS), which implements decoding. In this paper, the achievable rate region of such a network is studied (particularized for the 2-user case). We devise an iterative algorithm that solves the weighted sum-rate optimization, and derives the optimum compression codebooks at the BSs. The extension to more than two users is straightforward.**

I. Introduction

Inter-cell interference is one of the most limiting factors of current cellular networks. To overcome it, designers have resorted to orthogonal frequency channels, sectorized antennas and fractional frequency reuse. However, a more spectrally efficient solution has been recently proposed: coordinated cellular networks [1], [2]. They consist of single-frequency networks with base stations (BSs) cooperating in order to receive from the mobile terminals.

Preliminary results on the uplink capacity of coordinated networks consider all BSs connected via a lossless backhaul with unlimited capacity (see [3] and references therein). Accordingly, the uplink capacity region equals that of the MIMO multi-access channel, with a supra-receiver containing all the antennas of all cooperative BSs. Such an assumption is optimistic for current cellular networks. To deal with a more realistic backhaul constraint, two main approaches have been already proposed: *i*) *distributed decoding* [4], consisting on a LLR exchange across neighbor BSs. The decoding delay is its main problem. *ii*) *Quantization* [5]: the BSs quantize their observations and forward them to the decoding unit via the constrained backhaul. Its main drawback is that it does not take profit of the signal correlation between BSs/antennas.

In this paper, we study another approach for the network: *distributed compression* [6]. The cooperative BSs, upon receiving their signals, distributely compress them using a multi-source lossy compression code [7]. Then, using the constrained backhaul, they transmit their compressed vectors to the central unit (also a BS), which de-compresses them using its own received signal as side information. The central unit finally uses the reconstructed vectors, as well as its own signal, to decode the users' messages. The optimum compression of multiple, correlated, sources to be decompressed at single central unit with side information is still unknown. To the best of authors knowledge, the scheme that achieves the tightest rate-distortion region for the problem is Distributed Wyner-Ziv lossy compression [8]. Such an architecture is the direct extension of Berger-Tung coding to the decoding side information case [7], [9]. In turn, Berger-Tung can be thought as the counterpart, for continuous sources, of the Slepian-Wolf lossless coding [10]. Distributed Wyner-Ziv is the coding scheme proposed to be used in the network.

Our study focusses on the achievable rate region of a coordinated network with $N+1$ multi-antenna BSs. The first BS, denoted BS_0, is the central unit and implements the users' decoding. The rest, $\mathrm{BS}_1, \cdots, \mathrm{BS}_N$, are cooperative BSs which communicate with BS_0 using a lossless wired backhaul of aggregate rate R. Each BS has N_i receive antennas, $i = 0, \cdots, N$. In the network, we assume two multiple-antenna users, s_1 and s_2, transmitting pre-defined, Gaussian, space-time codewords over *time-invariant*, *frequency-flat* channels. Both users transmit simultaneously, in the same frequency band and with N_t antennas. Receive channel state information is assumed at the the decoding unit. Notice that our analysis complements the single-antenna results in [6].

The remainder of the paper is organized as follows: Sec. II briefly introduces distributed Wyner-Ziv compression. In Sec. III we state the problem and give an outer region on the 2-user's rate region. Sec. IV solves the weighted sum-rate optimization for the problem, using an iterative algorithm based upon dual decomposition and gradient projection. Finally, Sec. V depicts numerical results.

Notation. We compactly write $\boldsymbol{Y}_{1:N} = \{\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_N\}$, $\boldsymbol{Y}_{\mathcal{G}} = \{\boldsymbol{Y}_i | i \in \mathcal{G}\}$ and $\boldsymbol{Y}_n^c = \{\boldsymbol{Y}_i | i \neq n\}$. Block-diagonal matrices are defined as $\mathrm{diag}\left(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_n\right)$, with $\boldsymbol{A}_i$ square matrices. A sequence of vectors $\{\boldsymbol{Y}_i^t\}_{t=1}^n$ is compactly denoted by $\boldsymbol{Y}_i^n$. $\mathrm{coh}\left(\cdot\right)$ stands for convex hull. Finally, we define $\boldsymbol{R}_{\boldsymbol{X}|\boldsymbol{Y}} = \boldsymbol{E}\left\{\left(\boldsymbol{X} - \boldsymbol{E}\left\{\boldsymbol{X}|\boldsymbol{Y}\right\}\right)\left(\boldsymbol{X} - \boldsymbol{E}\left\{\boldsymbol{X}|\boldsymbol{Y}\right\}\right)^\dagger |\boldsymbol{Y}\right\}$.

II. Distributed Wyner-Ziv Compression

Let $\boldsymbol{Y}_i^n,\ i = 1, \cdots, N$ be N zero-mean, temporally memoryless, jointly Gaussian vectors to be compressed independently at $\mathrm{BS}_1, \cdots, \mathrm{BS}_N$, respectively. Assume that they are the observations at the BSs of the sum of signals transmitted by user s_1 and user s_2, *i.e.,* $\boldsymbol{X}_1^n$ and $\boldsymbol{X}_2^n$, respectively. The compressed vectors are sent to the central unit BS_0, which jointly de-compresses them using its received signal $\boldsymbol{Y}_0^n$ as side information, and recovers the BSs' observations as $\hat{\boldsymbol{Y}}_i^n,\ i=1,\cdots,N$. Distributed Wyner-Ziv compression [8] applies for such a setup as follows:

Definition 1 (Multiple-source Compression Code): A compression code $(n, 2^{n\rho_1}, \cdots, 2^{n\rho_N})$ with side information at the decoder $\boldsymbol{Y}_0$ is defined by $N+1$ mappings, $f_n^i(\cdot)$, $i=1,\cdots,N$, and $g_n(\cdot)$, and $2N+1$ spaces $\mathcal{Y}_i, \hat{\mathcal{Y}}_i$, $i=1,\cdots,N$ and $\mathcal{Y}_0$, where

$$f_n^i : \mathcal{Y}_i^n \rightarrow \{1, \cdots, 2^{n\rho_i}\}, \quad i=1,\cdots,N$$
$$g_n : \{1, \cdot, 2^{n\rho_1}\} \times \cdot \times \{1, \cdot, 2^{n\rho_N}\} \times \mathcal{Y}_0^n \rightarrow \hat{\mathcal{Y}}_1^n \times \cdot \times \hat{\mathcal{Y}}_N^n.$$

Proposition 1 (Distributed Wyner-Ziv [8]): Let the random vectors $\hat{\boldsymbol{Y}}_i$, $i=1,\cdots,N$ have conditional probability $p\left(\hat{\boldsymbol{Y}}_i|\boldsymbol{Y}_i\right)$ and satisfy the Markov chain $\left(\boldsymbol{Y}_0, \boldsymbol{Y}_i^c, \hat{\boldsymbol{Y}}_i^c\right) \rightarrow \boldsymbol{Y}_i \rightarrow \hat{\boldsymbol{Y}}_i$. Let also $\boldsymbol{Y}_0$ and $\boldsymbol{Y}_i$, $i=1,\cdots,N$ be jointly Gaussian. Then, considering a sequence of compression codes $(n, 2^{n\rho_1}, \cdots, 2^{n\rho_N})$ with side information $\boldsymbol{Y}_0$ at the decoder, the following mutual information satisfies

$$\frac{1}{n} I\left(\boldsymbol{X}_{\mathcal{U}}^n; \boldsymbol{Y}_0^n, g_n\left(\boldsymbol{Y}_0^n, f_n^1\left(\boldsymbol{Y}_1^n\right), \cdots, f_n^N\left(\boldsymbol{Y}_N^n\right)\right) | \boldsymbol{X}_{\mathcal{U}^c}^n\right) = I\left(\boldsymbol{X}_{\mathcal{U}}; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N} | \boldsymbol{X}_{\mathcal{U}^c}\right), \quad \forall \mathcal{U} \subseteq \{1,2\} \quad (1)$$

as $n \rightarrow \infty$ if:

- the compression rates $\rho_1, \cdots, \rho_N$ satisfy
$$I\left(\boldsymbol{Y}_{\mathcal{G}}; \hat{\boldsymbol{Y}}_{\mathcal{G}} | \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{\mathcal{G}}^c\right) \leq \sum_{i\in\mathcal{G}} \rho_i \qquad \forall \mathcal{G} \subseteq \{1,\cdots,N\}, \quad (2)$$
- each compression codebook C_i, $i=1,\cdots,N$ consists of $2^{n\rho_i}$ random sequences $\hat{\boldsymbol{Y}}_i^n$ drawn *i.i.d.* from $\prod_{t=1}^n p\left(\hat{\boldsymbol{Y}}_i\right)$, where $p\left(\hat{\boldsymbol{Y}}_i\right) = \sum_{\boldsymbol{Y}_i} p\left(\boldsymbol{Y}_i\right) p\left(\hat{\boldsymbol{Y}}_i|\boldsymbol{Y}_i\right)$.
- for every $i=1,\cdots,N$, the encoding $f_n^i(\cdot)$ outputs the bin-index of codewords $\hat{\boldsymbol{Y}}_i^n$ that are jointly typical with the source sequence $\boldsymbol{Y}_i^n$. In turn, $g_n(\cdot)$ outputs the codewords $\hat{\boldsymbol{Y}}_i^n$, $i=1,\cdots,N$ that, belonging to the bins selected by the encoders, are all jointly typical with $\boldsymbol{Y}_0^n$.

Proof: The statement is proven for discrete sources and discrete side information in [8, Theorem 2]. Also, the extension to the Gaussian case is conjectured therein. The conjecture can be proven by noting that Distributed Wyner-Ziv coding is equivalent to Berger-Tung coding with side information at the decoder. In turn, Berger-Tung coding can be implemented through time-sharing of successive Wyner-Ziv compressions [9], for which introducing side information $\boldsymbol{Y}_0$ at the decoder reduces the compression rate as in (2). Hence, the statement holds. Due to space limitations, we refer the reader to [11] for the complete proof. ■

III. Problem Statement

Let the two users transmit simultaneously two independent messages ω_1 and ω_2. The messages are mapped onto two zero-mean, Gaussian codewords $\boldsymbol{X}_1^n$ and $\boldsymbol{X}_2^n$, drawn *i.i.d.* from random vectors $\boldsymbol{X}_u \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{Q}_u)$, $u=1,2$, where $\boldsymbol{Q}_u$ are not subject to optimization. The transmitted signals, affected by *time-invariant*, *frequency-flat* fading, are received at the $N+1$ BSs under additive noise:

$$\boldsymbol{Y}_i^n = \sum_{u=1}^{2} \boldsymbol{H}_{u,i} \cdot \boldsymbol{X}_u^n + \boldsymbol{Z}_i^n, \quad i=0,\cdots,N, \qquad (3)$$

where $\boldsymbol{H}_{u,i}$ is the MIMO channel matrix between user u and BS_i, and $\boldsymbol{Z}_i \sim \mathcal{CN}\left(0, \sigma_r^2 \boldsymbol{I}\right)$ is AWGN. The cooperative BS_1,...,BS_N, upon receiving their signals, compress them using a Distributed Wyner-Ziv code, and forward them to BS_0. As mentioned, we assume an overall backhaul constraint for communicating to BS_0, but not individual constraints; that is $\sum_{i=1}^N \rho_i \leq \mathrm{R}$. Accordingly, the set of constraints in (2) are all embedded into the global constraint: $I\left(\boldsymbol{Y}_{1:N}; \hat{\boldsymbol{Y}}_{1:N} | \boldsymbol{Y}_0\right) \leq \mathrm{R}$ (see [11] for the proof).

A. Achievable Rate Region

The set $\mathcal{C}$ of transmission rates at which messages ω_u, $u=1,2$ can be reliably decoded equals[1]:

$$\mathcal{C} = \mathrm{coh}\left(\bigcup_{\substack{\prod_{i=1}^N p\left(\hat{\boldsymbol{Y}}_i|\boldsymbol{Y}_i\right): \\ I\left(\boldsymbol{Y}_{1:N}; \hat{\boldsymbol{Y}}_{1:N}|\boldsymbol{Y}_0\right) \leq \mathrm{R}}} \mathcal{C}\left(\hat{\boldsymbol{Y}}_{1:N}\right) \right) \qquad \text{where} \qquad (4)$$

$$\mathcal{C}\left(\hat{\boldsymbol{Y}}_{1:N}\right) = \left\{ (R_{1,2}) : \begin{array}{r} R_1 \leq I\left(\boldsymbol{X}_1; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N} | \boldsymbol{X}_2\right) \\ R_2 \leq I\left(\boldsymbol{X}_2; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N} | \boldsymbol{X}_1\right) \\ R_1 + R_2 \leq I\left(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N}\right) \end{array} \right\} \qquad (5)$$

Where (5) comes out directly from (1) in Prop. 1, and (4) by noting that compression codebooks can be arbitrary chosen at the BSs. Notice that the boundary points of the region can be achieved using superposition coding (SC) at the users, successive interference cancellation (SIC) at the BS_0, and (optionally) *time-sharing* (TS).

Equivalently to the single-user case (see [11, Proposition 1]), the optimum choice of $\hat{\boldsymbol{Y}}_i$, $i=1,\cdots,N$ at the boundary can be shown to be jointly Gaussian with $\boldsymbol{Y}_i$, $i=1,\cdots,N$. Therefore, the union in (4) can be restricted to the compression vectors satisfying $\hat{\boldsymbol{Y}}_i = \boldsymbol{Y}_i + \boldsymbol{Z}_i^c$; being $\boldsymbol{Z}_i^c \sim \mathcal{CN}(0, \boldsymbol{\Phi}_i)$ independent, Gaussian, "compression" noise at BS_i. That is, the rate region remains as in (6), where $c(\mathrm{R}) = \left\{\boldsymbol{\Phi}_{1:N} : \log\det\left(\boldsymbol{I} + \mathrm{diag}\left(\boldsymbol{\Phi}_1^{-1}, \cdots, \boldsymbol{\Phi}_N^{-1}\right) \boldsymbol{R}_{\boldsymbol{Y}_{1:N}|\boldsymbol{Y}_0}\right) \leq \mathrm{R}\right\}$, $\boldsymbol{Q} = \mathrm{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$ and $\boldsymbol{H}_{s,n} = [\boldsymbol{H}_{1,n}, \ \boldsymbol{H}_{2,n}]$, for $n=0,\cdots,N$. The conditional covariance $\boldsymbol{R}_{\boldsymbol{Y}_{1:N}|\boldsymbol{Y}_0}$ is computed as [11, Appendix 1] :

$$\boldsymbol{R}_{\boldsymbol{Y}_{1:N}|\boldsymbol{Y}_0} = \begin{bmatrix} \boldsymbol{H}_{s,1} \\ \vdots \\ \boldsymbol{H}_{s,N} \end{bmatrix} \left(\boldsymbol{I} + \frac{\boldsymbol{Q}}{\sigma_r^2} \boldsymbol{H}_{s,0}^\dagger \boldsymbol{H}_{s,0}\right)^{-1} \boldsymbol{Q} \begin{bmatrix} \boldsymbol{H}_{s,1} \\ \vdots \\ \boldsymbol{H}_{s,N} \end{bmatrix}^\dagger + \sigma_r^2 \boldsymbol{I}.$$

To compute the boundary points of this rate region, we resort to the definition given by Cheng and Verdú in [12, Section III-C], where bounding hyperplanes are used to describe it:

$$\mathcal{C} = \left\{(R_{1,2}) : \alpha R_1 + (1-\alpha) R_2 \leq \mathcal{R}(\alpha), \forall \alpha \in [0,1]\right\}. \quad (7)$$

[1] As mentioned, users' covariance are fixed and not subject to optimization.

$$\mathcal{C} = \operatorname{coh}\left(\bigcup_{\substack{\mathbf{\Phi}_1,\cdots,\mathbf{\Phi}_N \\ \in c(\mathrm{R})}} \left\{ \begin{array}{c} R_1 \leq \log\det\left(\boldsymbol{I} + \frac{\boldsymbol{Q}_1}{\sigma_r^2}\boldsymbol{H}_{1,0}^{\dagger}\boldsymbol{H}_{1,0} + \boldsymbol{Q}_1 \sum_{n=1}^{N} \boldsymbol{H}_{1,n}^{\dagger}\left(\sigma_r^2\boldsymbol{I} + \mathbf{\Phi}_n\right)^{-1}\boldsymbol{H}_{1,n}\right) \\ R_2 \leq \log\det\left(\boldsymbol{I} + \frac{\boldsymbol{Q}_2}{\sigma_r^2}\boldsymbol{H}_{2,0}^{\dagger}\boldsymbol{H}_{2,0} + \boldsymbol{Q}_2 \sum_{n=1}^{N} \boldsymbol{H}_{2,n}^{\dagger}\left(\sigma_r^2\boldsymbol{I} + \mathbf{\Phi}_n\right)^{-1}\boldsymbol{H}_{2,n}\right) \\ R_1 + R_2 \leq \log\det\left(\boldsymbol{I} + \frac{\boldsymbol{Q}}{\sigma_r^2}\boldsymbol{H}_{s,0}^{\dagger}\boldsymbol{H}_{s,0} + \boldsymbol{Q} \sum_{n=1}^{N} \boldsymbol{H}_{s,n}^{\dagger}\left(\sigma_r^2\boldsymbol{I} + \mathbf{\Phi}_n\right)^{-1}\boldsymbol{H}_{s,n}\right) \end{array} \right\} \right) \quad (6)$$

$\mathcal{R}(\alpha)$ is the maximum weighted sum-rate (WSR) of the network, considering the weights α and $(1-\alpha)$ for user s_1 and s_2, respectively. This WSR is achieved with equality at the boundary of the rate region [12], whose points can be attained considering SIC at BS_0 (as previously mentioned). We solve the WSR in Sec. IV. First, we present two useful upper bounds on the rate region.

B. Outer Regions

Outer Bound 1: The achievable rate region (4) is contained within the region

$$\begin{aligned} R_1 &\leq \log\det\left(\boldsymbol{I} + \frac{\boldsymbol{Q}_1}{\sigma_r^2}\sum_{n=0}^{N}\boldsymbol{H}_{1,n}^{\dagger}\boldsymbol{H}_{1,n}\right) \\ R_2 &\leq \log\det\left(\boldsymbol{I} + \frac{\boldsymbol{Q}_2}{\sigma_r^2}\sum_{n=0}^{N}\boldsymbol{H}_{2,n}^{\dagger}\boldsymbol{H}_{2,n}\right) \\ R_1 + R_2 &\leq \log\det\left(\boldsymbol{I} + \frac{\boldsymbol{Q}}{\sigma_r^2}\sum_{n=0}^{N}\boldsymbol{H}_{s,n}^{\dagger}\boldsymbol{H}_{s,n}\right) \end{aligned} \quad (8)$$

where $\boldsymbol{Q} = \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$ and $\boldsymbol{H}_{s,n} = [\boldsymbol{H}_{1,n},\ \boldsymbol{H}_{2,n}]$, for $n = 0, \cdots, N$.

Remark 1: It is the capacity rate region of the system when $\boldsymbol{Y}_i$, $i = 1, \cdots, N$ are available at BS_0 directly without compression.

Outer Bound 2: The following relationship between the sum-rate $R_1 + R_2$ and the backhaul rate R holds

$$R_1 + R_2 \leq \log\det\left(\boldsymbol{I} + \frac{1}{\sigma_r^2}\boldsymbol{H}_{s,0}\boldsymbol{Q}\boldsymbol{H}_{s,0}^{\dagger}\right) + \mathrm{R}. \quad (9)$$

where $\boldsymbol{Q} = \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$ and $\boldsymbol{H}_{s,n} = [\boldsymbol{H}_{1,n},\ \boldsymbol{H}_{2,n}]$, for $n = 1, \cdots, N$.

Proof: See [11, Appendix III] for the proof. ■

IV. Weighted Sum-Rate Optimization

Let solve the WSR optimization for $\alpha \geq \frac{1}{2}$ (*i.e.,* higher priority to user 1, which is decoded last at the SIC). With such a scheme, the maximum transmission rate of user 1 follows

$$\begin{aligned} R_1 &= I\left(\boldsymbol{X}_1; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N} | \boldsymbol{X}_2\right) \quad (10) \\ &= \log\det\left(\boldsymbol{I} + \frac{\boldsymbol{Q}_1}{\sigma_r^2}\boldsymbol{H}_{1,0}^{\dagger}\boldsymbol{H}_{1,0}\right. \\ &\quad \left. + \boldsymbol{Q}_1\sum_{n=1}^{N}\boldsymbol{H}_{1,n}^{\dagger}\left(\sigma_r^2\boldsymbol{I} + \mathbf{\Phi}_n\right)^{-1}\boldsymbol{H}_{1,n}\right). \end{aligned}$$

On the other hand, the rate of user 2, which is decoded first, follows:

$$\begin{aligned} R_2 &= I\left(\boldsymbol{X}_2; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N}\right) \quad (11) \\ &= I\left(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N}\right) - I\left(\boldsymbol{X}_1; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N} | \boldsymbol{X}_2\right) \\ &= \log\det\left(\boldsymbol{I} + \frac{\boldsymbol{Q}}{\sigma_r^2}\boldsymbol{H}_{s,0}^{\dagger}\boldsymbol{H}_{s,0} + \right. \\ &\quad \left. \boldsymbol{Q}\sum_{n=1}^{N}\boldsymbol{H}_{s,n}^{\dagger}\left(\sigma_r^2\boldsymbol{I} + \mathbf{\Phi}_n\right)^{-1}\boldsymbol{H}_{s,n}\right) - R_1, \end{aligned}$$

where $\boldsymbol{Q} = \operatorname{diag}(\boldsymbol{Q}_1, \boldsymbol{Q}_2)$ and $\boldsymbol{H}_{s,n} = [\boldsymbol{H}_{1,n},\ \boldsymbol{H}_{2,n}]$. The WSR, $\alpha R_1 + (1-\alpha)R_2$, which has to be maximized is convex on the compression noises $\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_N$. To make the optimization concave, we introduce the change of variables $\mathbf{\Phi}_n = \boldsymbol{A}_n^{-1}$, $n = 1, \cdots, N$. Considering so, and introducing (10) and (11) in (7), the WSR optimization remains

$$\begin{aligned} \mathcal{R}(\alpha) = &\max_{\boldsymbol{A}_1,\cdots,\boldsymbol{A}_N} \ \alpha \cdot R_1 + (1-\alpha) \cdot R_2 \quad (12) \\ &\text{s.t.} \quad \log\det\left(\boldsymbol{I} + \operatorname{diag}\left(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N\right)\boldsymbol{R}_{\boldsymbol{Y}_{1:N}|\boldsymbol{Y}_0}\right) \leq \mathrm{R} \end{aligned}$$

Although the objective function has turned into concave on $\boldsymbol{A}_n \succeq 0$, $n = 1, \cdots, N$, the constraint now does not define a feasible convex set. Hence, the optimization is not convex in standard form. Our strategy to solve such an optimization is the following: first, we show that the optimization has zero duality gap. Later, we propose an iterative algorithm that solves the dual problem, thus solving the primal too.

Lemma 1: The duality gap for the WSR optimization (12) is zero.

Proof: Applying the time-sharing property in [13, Theorem 1] the zero-duality gap is demonstrated. See [11, Lemma 1] for the complete proof. ■

A. The Dual Problem

Let then solve the dual problem using an iterative algorithm. The Lagrangian for the WSR optimization is defined on $\lambda \geq 0$ and $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_n \succeq 0$ as:

$$\begin{aligned} &\mathcal{L}_\alpha\left(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_n, \lambda\right) = \alpha \cdot R_1 + (1-\alpha) \cdot R_2 \quad (13) \\ &-\lambda \cdot \left(\log\det\left(\boldsymbol{I} + \operatorname{diag}\left(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N\right)\boldsymbol{R}_{\boldsymbol{Y}_{1:N}|\boldsymbol{Y}_0}\right) - \mathrm{R}\right) \end{aligned}$$

The first step is to find the dual function [14, Sec. 5]

$$g_\alpha(\lambda) = \max_{\boldsymbol{A}_1,\cdots,\boldsymbol{A}_n \succeq 0} \mathcal{L}_\alpha\left(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_n, \lambda\right) \quad (14)$$

In order to solve (14), we propose the use of an iterative algorithm: the *gradient projection method* (GP) [14, Section 2.3]. It iterates as follows: let the maximization (14) and consider the initial point $\left\{\boldsymbol{A}_1^0, \cdots, \boldsymbol{A}_n^0\right\} \succeq 0$. We update the values as [14, Section 2.3.1]

$$\boldsymbol{A}_n^{t+1} = \boldsymbol{A}_n^t + \gamma_t\left(\bar{\boldsymbol{A}}_n^t - \boldsymbol{A}_n^t\right), \quad n = 1, \cdots, N \quad (15)$$

where t is the iteration index, $0 < \gamma_t \leq 1$ is the step size, and

$$\bar{\boldsymbol{A}}_n^t = \left[\boldsymbol{A}_n^t + s_t \cdot \nabla_{\boldsymbol{A}_n}\mathcal{L}_\alpha\left(\lambda, \boldsymbol{A}_1^t, \cdots, \boldsymbol{A}_N^t\right)\right]_{\succeq 0}, \quad n = 1, \cdot, N \quad (16)$$

with $s_t \geq 0$ a scalar and $\nabla_{\boldsymbol{A}_n}\mathcal{L}_\alpha\left(\lambda, \boldsymbol{A}_1^t, \cdots, \boldsymbol{A}_N^t\right)$ the gradient of $\mathcal{L}_\alpha(\cdot)$ with respect to $\boldsymbol{A}_n$, evaluated at $\boldsymbol{A}_1^t, \cdots, \boldsymbol{A}_N^t$. Finally, $[\cdot]_{\succeq 0}$ denotes the projection onto the cone of positive semidefinite matrices. Provided that γ_t and s_t are chosen appropriately, the sequence $\{\boldsymbol{A}_1^t, \cdots, \boldsymbol{A}_n^t\}$ is guaranteed to converge to a local maximum of (14) [14, Proposition 2.2.1]. To demonstrate convergence to the global maximum, and therefore to $g_\alpha(\lambda)$, it is necessary to prove that the mapping

$T(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N) = [\boldsymbol{A}_1 + \gamma \nabla_{\boldsymbol{A}_1} \mathcal{L}_\alpha, \cdots, \boldsymbol{A}_N + \gamma \nabla_{\boldsymbol{A}_N} \mathcal{L}_\alpha]$ is a block contraction[2] for some γ [15, Proposition 3.10]. We were not able to mathematically demonstrate the contraction property on the Lagrangian. However, simulation results show convergence of our algorithm to the global maximum always.

To make the algorithm work for the problem, we need to: *i*) compute the projection of a Hermitian matrix $\boldsymbol{S}$, with eigen-decomposition $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\eta}\boldsymbol{U}^\dagger$, onto the cone of positive semidefinite matrices, which is equal to [16, Theorem 2.1]:

$$[\boldsymbol{S}]_{\succeq 0} = \boldsymbol{U} \text{diag}\left(\max\{\eta_1, 0\}, \cdots, \max\{\eta_m, 0\}\right) \boldsymbol{U}^\dagger. \quad (19)$$

ii) Obtain the gradient of $\mathcal{L}_\alpha(\cdot)$ with respect to a single $\boldsymbol{A}_n$. First, we recall that $\boldsymbol{A}_n$, $n = 1, \cdots, N$ are complex matrices and the lagrangian is real-valued. Hence, the gradient of the function with respect to $\boldsymbol{A}_n$ is equal to twice the conjugate of the partial derivative of the function with respect to such a matrix [17]:

$$\nabla_{\boldsymbol{A}_n} \mathcal{L}_\alpha(\boldsymbol{A}_{1:N}, \lambda) = 2\left(\left[\frac{\partial \mathcal{L}_\alpha(\boldsymbol{A}_{1:N}, \lambda)}{\partial \boldsymbol{A}_n}\right]^T\right)^\dagger$$

The Lagrangian is defined in (13). To obtain its partial derivative, we first make use of results in [11, Appendix V] to derive:

$$\left[\frac{\partial \log\det\left(\boldsymbol{I} + \text{diag}(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N) \boldsymbol{R}_{\boldsymbol{Y}_{1:N}|\boldsymbol{Y}_0}\right)}{\partial \boldsymbol{A}_n}\right]^T \quad (20)$$
$$= \left[\frac{\partial \log\det\left(\boldsymbol{I} + \boldsymbol{A}_n \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c}\right)}{\partial \boldsymbol{A}_n}\right]^T$$
$$= \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c}\left(\boldsymbol{I} + \boldsymbol{A}_n \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c}\right)^{-1}$$

where the conditional covariance is computed in (17). We can also derive that

$$\frac{\partial R_1}{\partial \boldsymbol{A}_n} = \frac{\partial I\left(\boldsymbol{X}_1; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N} | \boldsymbol{X}_2\right)}{\partial \boldsymbol{A}_n} \quad (21)$$
$$= \frac{\partial I\left(\boldsymbol{X}_1; \hat{\boldsymbol{Y}}_n | \boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c\right)}{\partial \boldsymbol{A}_n}$$

where second equality follows from the chain rule for mutual information and noting that $I\left(\boldsymbol{X}_1; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c | \boldsymbol{X}_2\right)$ does not depend on $\boldsymbol{A}_n$. The mutual information above is computed as follows:

$$I\left(\boldsymbol{X}_1; \hat{\boldsymbol{Y}}_n | \boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c\right) = H\left(\hat{\boldsymbol{Y}}_n | \boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c\right) \quad (22)$$
$$- H\left(\hat{\boldsymbol{Y}}_n | \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c\right)$$
$$= \log\det\left(\boldsymbol{A}_n \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c} + \boldsymbol{I}\right)$$
$$- \log\det\left(\boldsymbol{A}_n \sigma_r^2 + \boldsymbol{I}\right)$$

The covariance $\boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c}$ is derived from (18). Therefore, the derivative of R_1 remains [17]:

$$\left[\frac{\partial R_1}{\partial \boldsymbol{A}_n}\right]^T = \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c}\left(\boldsymbol{A}_n \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c} + \boldsymbol{I}\right)^{-1} \quad (23)$$
$$- \sigma_r^2 \left(\boldsymbol{A}_n \sigma_r^2 + \boldsymbol{I}\right)^{-1}.$$

[2]See [15, Section 3.1.2] for the definition of block-contraction.

Equivalently, we can derive for the derivative of R_2 that

$$\frac{\partial R_2}{\partial \boldsymbol{A}_n} = \frac{\partial I\left(\boldsymbol{X}_2; \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_{1:N}\right)}{\partial \boldsymbol{A}_n} \quad (24)$$
$$= \frac{\partial I\left(\boldsymbol{X}_2; \hat{\boldsymbol{Y}}_n | \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c\right)}{\partial \boldsymbol{A}_n}.$$

Such a mutual information is computed as:

$$I\left(\boldsymbol{X}_2; \hat{\boldsymbol{Y}}_n | \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c\right) = H\left(\hat{\boldsymbol{Y}}_n | \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c\right) - H\left(\hat{\boldsymbol{Y}}_n | \boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c\right) \quad (25)$$
$$= \log\det\left(\boldsymbol{A}_n \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c} + \boldsymbol{I}\right)$$
$$- \log\det\left(\boldsymbol{A}_n \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c} + \boldsymbol{I}\right)$$

where conditional covariances are obtained in (17) and (18), respectively. Therefore, the derivative of R_2 is:

$$\left[\frac{\partial R_2}{\partial \boldsymbol{A}_n}\right]^T = \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c}\left(\boldsymbol{A}_n \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c} + \boldsymbol{I}\right)^{-1} \quad (26)$$
$$- \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c}\left(\boldsymbol{A}_n \boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{X}_2, \boldsymbol{Y}_0, \hat{\boldsymbol{Y}}_n^c} + \boldsymbol{I}\right)^{-1}.$$

Plugging (20), (23) and (26) into (20) we obtain the gradient of the function. Notice that for $\alpha \leq \frac{1}{2}$, the roles of users s_1 and s_2 are interchanged, being user 1 decoded first. The roles would also need to be interchanged in the computation of the gradients of R_1 and R_2. Once obtained the dual function through GP, we minimize it to obtain:

$$\mathcal{R}(\alpha) = \min_{\lambda \geq 0} \ g_\alpha(\lambda). \quad (27)$$

To solve this minimization, we use the subgradient approach as in [18, Algorithm 1], which consists on following search direction $-h$, where

$$h = \mathrm{R} - \log\det\left(\boldsymbol{I} + \text{diag}(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N) \boldsymbol{R}_{\boldsymbol{Y}_{1:N}|\boldsymbol{Y}_0}\right) \quad (28)$$

Taking all this into account we build up the Algorithm 1 to solve the dual problem of the WSR, and thus the primal too. We can only claim local convergence of the algorithm, even though simulation results suggest convergence to the global maximum always.

Algorithm 1 Two-user WSR Dual Problem

Initialize $\lambda_{\min} = 0$ and $\lambda_{\max}$
repeat
  $\lambda = \frac{\lambda_{\max} - \lambda_{\min}}{2}$
  Obtain $\{\boldsymbol{A}_1^*, \cdots, \boldsymbol{A}_N^*\} = \arg\max \mathcal{L}_\alpha(\boldsymbol{A}_{1:N}, \lambda)$ from Algorithm 2
  Evaluate h as in (28).
  if $h \leq 0$, then $\lambda_{\min} = \lambda$, else $\lambda_{\max} = \lambda$
until $\lambda_{\max} - \lambda_{\min} \leq \epsilon$
$\mathcal{R}(\alpha) = \alpha R_1(\boldsymbol{A}_{1:N}^*) + (1-\alpha) R_2(\boldsymbol{A}_{1:N}^*)$.

$$\boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{Y}_0,\hat{\boldsymbol{Y}}_n^c} = \boldsymbol{H}_{s,n}\left(\boldsymbol{I}+\boldsymbol{Q}\left(\frac{1}{\sigma_r^2}\boldsymbol{H}_{s,0}^\dagger\boldsymbol{H}_{s,0}+\sum_{j\neq n}\boldsymbol{H}_{s,j}^\dagger\left(\boldsymbol{A}_j\sigma_r^2\boldsymbol{I}+\boldsymbol{I}\right)^{-1}\boldsymbol{A}_j\boldsymbol{H}_{s,j}\right)\right)^{-1}\boldsymbol{Q}\boldsymbol{H}_{s,n}^\dagger+\sigma_r^2\boldsymbol{I} \quad (17)$$

$$\boldsymbol{R}_{\boldsymbol{Y}_n|\boldsymbol{X}_2,\boldsymbol{Y}_0,\hat{\boldsymbol{Y}}_n^c} = \boldsymbol{H}_{1,n}\left(\boldsymbol{I}+\frac{\boldsymbol{Q}_1}{\sigma_r^2}\boldsymbol{H}_{1,0}^\dagger\boldsymbol{H}_{1,0}+\sum_{p\neq n}\boldsymbol{Q}_1\boldsymbol{H}_{1,p}^\dagger\left(\boldsymbol{A}_p\sigma_r^2\boldsymbol{I}+\boldsymbol{I}\right)^{-1}\boldsymbol{A}_p\boldsymbol{H}_{1,p}\right)^{-1}\boldsymbol{Q}_1\boldsymbol{H}_{1,n}^\dagger+\sigma_r^2\boldsymbol{I} \quad (18)$$

Algorithm 2 GP to obtain $g_\alpha(\lambda)$

1: Initialize $\boldsymbol{A}_n^0 = \boldsymbol{0}$, $n = 1, \cdots, N$ and $t = 0$
2: **repeat**
3: Compute the gradient $\boldsymbol{G}_n^t = \nabla_{\boldsymbol{A}_n}\mathcal{L}_\alpha(\lambda, \boldsymbol{A}_1^t, \cdots, \boldsymbol{A}_N^t)$, $n = 1, \cdots, N$ from (20).
4: Choose appropriate s_t
5: Set $\hat{\boldsymbol{A}}_n^t = \boldsymbol{A}_n^t + s_t \cdot \boldsymbol{G}_n^t$, $n = 1, \cdots, N$, compute $\hat{\boldsymbol{A}}_n^t = \boldsymbol{U}\boldsymbol{\eta}\boldsymbol{U}^\dagger$, and project $\bar{\boldsymbol{A}}_n^t = \boldsymbol{U}\max\{\boldsymbol{\eta}, 0\}\boldsymbol{U}^\dagger$.
6: Choose appropriate γ_t
7: Update $\boldsymbol{A}_n^{t+1} = \boldsymbol{A}_n^t + \gamma_t\left(\bar{\boldsymbol{A}}_n^t - \boldsymbol{A}_n^t\right)$, $n = 1, \cdots, N$
8: $t = t + 1$
9: **until** The sequence converges $\{\boldsymbol{A}_1^t, \cdots, \boldsymbol{A}_N^t\} \rightarrow \{\boldsymbol{A}_1^*, \cdots, \boldsymbol{A}_N^*\}$
10: Return $\{\boldsymbol{A}_1^*, \cdots, \boldsymbol{A}_N^*\}$

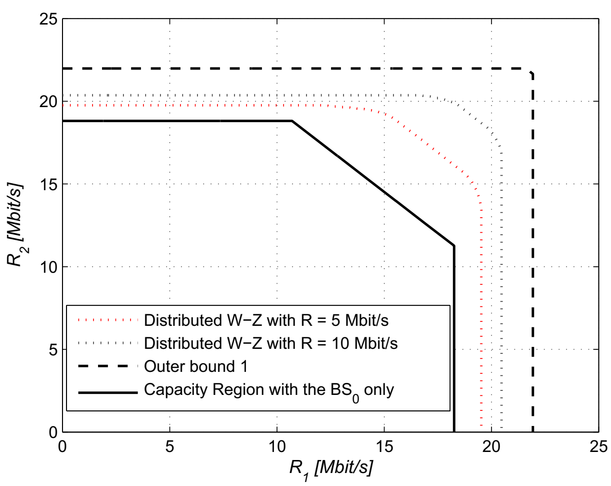

Fig. 1. The 2-user Rate Region, for different values of the backhaul rate R.

V. Numerical Results

Fig. 1 depicts the uplink rate region of a cellular network composed of BS_0 and its first tier of six cells. The radius of each cell is 700 m, and all BSs have 3 antennas. Within the network, wireless channels are simulated with path loss, shadowing and *i.i.d* Rayleigh fading among antennas. We consider Line-of-Sight (LoS) propagation, with path-loss exponent $\alpha = 2.6$, and shadowing standard deviation $\sigma = 4$ dB. The transmission bandwidth is set to 1 MHz and the carrier frequency is 2.5 GHz. The two users are equipped with 2 TX antennas each, and placed at the edge of the central cell. Finally, the transmitters' power is 23 dBm, with isotropic transmission, *i.e.*, $\boldsymbol{Q}_u = \frac{P}{2}\boldsymbol{I}$, $u = 1, 2$.

It is clearly shown that the region is significantly enlarged with only 5 Mbit/s of backhaul rate, shared among 6 cooperative BSs. Also, there is a definite advantage on sum-rate for the network.

References

[1] A.D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. on Information Theory*, vol. 40, pp. 17131727, Nov. 1994.
[2] G.J. Foschini, K. Karakayali, and R.A. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," *IEE Proceedings Communications*, vol. 153, no. 4, pp. 548–555, Aug. 2006.
[3] O. Somekh, O. Simeone, Y. Bar-ness, A. Haimovich, U. Spagnolini, and S. Shamai, *An Information Theoretic view of distributed antenna processing in cellular systems*, Auerbach Publication, CRC Press, 2007.
[4] E. Aktas, J. Evans, and S. Hanly, "Distributed decoding in a cellular multiple-access channel," in *Proc. IEEE International Symposium on Infomation Theory*, Chicago, IL, Jun. 2004, p. 484.
[5] P. Marsch and G. Fettweis, "A framework for optimizing the uplink performnce of distributed antenna systems under a constrained backhaul," in *Proc. IEEE International Conference on Communications (ICC)*, Glasgow, UK, Jun. 2007.
[6] A. Sanderovich, O. Somekh, and S. Shamai (Shitz), "Uplink macro diversity with limited backhaul capacity," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Nice, France, Jun. 2007.
[7] S.Y. Tung, *Multiterminal source coding*, PhD Dissertation, Cornell University, 1978.
[8] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. on Information Theory*, vol. 50, no. 11, Nov. 2004.
[9] J. Chen and T. Berger, "Successive Wyner-Ziv coding scheme and its implications to the quadratic Gaussian CEO problem," *submitted IEEE Trans. on Information Theory*, 2006.
[10] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471–481, Jul. 1973.
[11] A. del Coso and S. Simoens, "Distributed compression for the uplink of a backhaul-constrained coordinated cellular network," *submitted to IEEE Trans. on Signal Processing*, 2008. arXiv:0802.0776.
[12] R.G. Cheng and S. Verdú, "Gaussian multiple-access channels with ISI: Capacity region and multi-user water-filling," *IEEE Trans. on Information Theory*, vol. 39, no. 3, pp. 773–785, May 1993.
[13] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. on Communications*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.
[14] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.
[15] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, Belmont, MA, 1997.
[16] J. Malick and H.S. Sendov, "Clarke generalized jacobian of the projection onto the cone of positive semidefinite matrices," *Springer Set-Valued Analysis*, vol. 14, no. 3, pp. 273–293, Sep. 2006.
[17] K.B. Petersen and M.S. Pedersen, *The Matrix Cookbook*, 2007.
[18] W. Yu, "A dual decomposition approach to the sum power Gaussian vector multiple-access channel sum capacity problem," in *Proc. Conference on Information Sciences and Systems*, The Johns Hopkins University, Mar. 2003.