Alternating Information Bottleneck Optimization for Weighted Sum Rate and Resource Allocation in the Uplink of C-RAN

Di Chen
Institute of Communications Engineering
University of Rostock
Email: di.chen@uni-rostock.de

Volker Kuehn
Institute of Communications Engineering
University of Rostock
Email: volker.kuehn@uni-rostock.de

***Abstract*—The compression step in the uplink of Cloud Radio Access Networks (C-RAN) and the resource allocation is studied in this paper. In C-RAN, there are multiple Radio Units (RUs) compressing and forwarding the correlated received signals simultaneously to the Central Processor (CP) in the cloud, via the fronthaul links with finite capacities. Due to the correlations, Wyner-Ziv coding is used. Hence a joint optimization of the quantizers is required at the CP. In this paper we aim to maximize the achievable weighted sum rate by optimizing the quantizers as well as the allocation of sum capacity to fronthaul links. At first we extend the Information Bottleneck (IB) method for the joint optimization of the quantizers used for compression. It is a combination of the Alternating Information Bottleneck method (AIB) and the Alternating Bi-Section method, which are both proposed in this paper. Then we use them to optimize the allocation of sum capacity.**

I. Introduction

(This is just a short version of introduction. The details will be available in the full version of the paper.)

It has been shown that a key feature of C-RAN is the transfer of baseband information to the CP via the capacity-constrained fronthaul links. Thus suitable compression strategies have to be developed in order to alleviate the requirements on the fronthaul links. There has been already some papers considering the optimization of quantization noise levels, when Compress and Forward (CF) or Noisy Network Coding (NNC) is performed at RUs, e.g. [6] and [7]. While these works consider only Gaussian codebooks, and treat the quantization as Gaussian test channels, the compression is modeled by adding Gaussian distributed quantization noise. Optimization of the quantization noise levels evaluates the performance only from the information theoretic perspective. In practice, the users might use arbitrary codebook $\mathcal{X}$ with finite alphabet, and the received signal is discretized and sampled firstly into finite alphabet $\mathcal{Y}$, then based on the compression scheme $P_{\hat{Y}|Y}$, it will be compressed into several quantization levels, denoted by $\hat{\mathcal{Y}}$. Usually $|\hat{\mathcal{Y}}|$ is much smaller than $|\mathcal{Y}|$ due to the compression. In such scenario, the Information Bottleneck (IB) method [8] is often used to optimize the quantizer $P_{\hat{Y}|Y}$ in order to maximize the objective mutual information.

However the IB method is considered only for the single quantizer case in most works. In C-RAN, where multiple quantizers exist, the optimization of the quantizers depends not only on its own channel configuration, but also on the other quantizers', since Wyner-Ziv coding is performed at the relays (CF) or joint decompression and decoding is performed at the destinations (NNC). Thus a joint optimization is required. The CP with high computing capability in the C-RAN makes this joint optimization to be possible. Then the question is whether the IB method can still be used. In this paper, we answer this question in the affirmative. We aim to maximize the achievable **weighted** sum rate in the uplink of C-RAN, where CF is performed at RUs. We propose an Alternating Information Bottleneck (AIB) method and an Alternating Bi-Section Method in order to achieve this goal.

It should be noted that the proposed AIB method and Alternating Bi-Section Method is applicable only to the scenario where the capacity of each fronthaul link is predetermined. This may happen when the fronthaul are optical fiber links. While the fronthaul might also be wireless, such that RUs share the total capacity resource. In this case the allocation of resource also influences the performance of the network. Based on the proposed algorithms as well as the Outer Linearization Method (OLM) [11], we propose an optimization scheme for the resource allocation.

Part of the results have been submitted to ICC 2016, where AIB method and Alternating-Bisection method are proposed. This paper is an extension, where weighted sum rate instead of sum rate, and the optimization of capacity allocation for different objectives are considered.

The remainder of the paper is organized as follows: In Sec. II we introduce the channel model considered and state the problem mathematically. Our optimization algorithms of the quantizers are presented and explained in Sec. III. The optimization algorithm of capacity allocation is shown in Sec. IV. Simulation results and conclusions are provided in Sec. V and Sec. VI respectively.

II. System Model and Problem Statement

A. System Model

We consider the C-RAN model depicted in Fig. 1. L single-antenna Mobile Users (MSs) send independent messages to L single-antenna RUs. The RUs are connected to a CP in the cloud via fronthaul links with finite sum capacity denoted by C_{sum} All messages need to be decoded at the CP. For

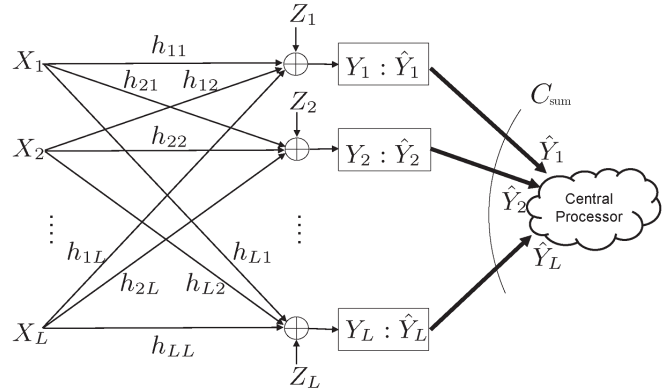

Fig. 1. The uplink of C-RAN with finite sum capacity fronthaul links [6].

simplicity, we consider only single antenna, but the algorithm can be extended to the MIMO case. Thus the channel model between the MSs and RUs is actually an $L \times L$ interference channel. The received analog signal at the i-th BS is

$$Y_{i,\text{analog}} = \sum_{j=1}^{L} h_{ij} X_j + Z_i,\ i \in \{1, 2, ..L\},$$

where $Z_i \sim \mathcal{CN}(0, \sigma_n^2)$ is the independent Gaussian noise with variance σ_n^2, and h_{ij} denotes the complex channel coefficient from the j-th MS to the i-th RU. X_i denotes the transmitted signal of the i-th MS, it can use arbitrary modulation scheme with available power $P_i = \mathbb{E}\{|X_i|^2\}$. The received analog signal $Y_{i,\text{analog}}$ is firstly sampled and discretized[1] into Y_i with finite alphabets $\mathcal{Y}_i$. Then each RU performs CF: Its quantizer compresses the signal Y_i into $\hat{Y}_i$ based on the compression scheme $P_{\hat{Y}_i|Y_i}$. $|\hat{\mathcal{Y}}_i|$ is assumed to be much smaller than $|\mathcal{Y}_i|$. Since the received signals of the neighboring RUs are statistically correlated, Wyner-Ziv coding is to be utilized. Then RUs independently send compressed bits to the CP via the fronthaul links. The CP employs successive two-stage decoding: It first decodes all the compressed signals $\hat{\boldsymbol{Y}} = [\hat{Y}_1, \hat{Y}_2, ...\hat{Y}_L]^T$ and then decodes MSs' messages $\boldsymbol{X} = [X_1, X_2, ..., X_L]^T$ based on the decoded compressed signals. Compared to NNC, where simultaneous joint decoding of compressed signals and the desired messages over all received blocks is required, the successive decoding nature of CF overcomes some difficulties in the practical implementation of NNC, such as long delay and high computational complexity. Moreover, we assume that the modulation scheme of each MS and CSI are known to the CP, and the design of the optimized quantizers can be feed-backed to the corresponding RU.

[1] This discretization is actually a pre-quantization, then the discretized signal should be further compressed by the quantizer due to limited fronthaul capacity. In this paper we address the optimization of the quantizer used for the compression.

B. Problem Statement

We aim to maximize the achievable weighted sum rate [2] in the uplink of C-RAN as follows.

$$\begin{aligned} &\max_{P_{\hat{\boldsymbol{Y}}|\boldsymbol{Y}}} \sum_{j=1}^{L} w_j R_j \\ &\text{Subject to } I(\boldsymbol{Y}; \hat{\boldsymbol{Y}}) \le C_{\text{sum}}, \end{aligned} \tag{1}$$

where $P_{\hat{\boldsymbol{Y}}|\boldsymbol{Y}} = \prod_{i=1}^{L} P_{\hat{Y}_i|Y_i}$. R_j and w_j denote the achievable rate of j-th MS and its weight, respectively. Since from CP's perspective, the network is actually MIMO-MAC, the capacity-achieving strategy in the MIMO-MAC is based on successive interference cancellation (SIC). Moreover, according to [12], the solution of (1) is given by the decoding order π that sorts the weights in non-increasing order

$$w_{\pi_1} \ge w_{\pi_2} \ge \cdots \ge w_{\pi_L}.$$

With this decoding order, the resulting maximization problem is convex. Since the decoding order is fixed solely by the weights, without loss of generality, we assume $w_L \ge w_{L-1} \ge \cdots \ge w_1$, i.e., x_1 is decoded first and x_L is decoded last. Thus we have

$$R_j = I(X_j; \hat{\boldsymbol{Y}} | X_1, X_2, ..., X_{j-1})\ \forall j \in \{1, 2, ..., L\} \tag{2}$$

The constraint of (1) can also be expressed as

$$\begin{aligned} &I(Y_i; \hat{Y}_i | \hat{Y}_1, \hat{Y}_2, ..., \hat{Y}_{i-1}) \le C_i\ \forall i \in \{1, 2, ..., L\}, \\ &\sum_{i=1}^{L} C_i = C_{\text{sum}}. \end{aligned} \tag{3}$$

where C_i denote the capacity allocated to i-th RU. It can be predetermined or optimized by the CP. We see that when the modulations schemes, the capacity of each fronthaul link and all channel configurations are fixed, the weighted sum rate depends only on how RUs compress their received signals. While when the capacity of each fronthaul link is not fixed, a simultaneous optimization of all quantizers and capacity allocation is necessary. In the following sections, we firstly assume that the capacity of each fronthaul link is fixed, and propose algorithms to optimize all quantizers jointly. Then based on the proposed algorithms, we propose an algorithm for the optimization of capacity allocation.

III. Optimization Algorithm and Quantizer Design (Fronthaul capacities predetermined)

In this section we assume the capacity of each fronthaul link is predetermined. Note that the fronthaul capacities are finite. On one hand, in the compression step, the quantization cannot be too fine, such that the compression rate exceeds the frontfaul capacity, in that case, the compressed information cannot be decoded at the CP. On the other hand, when the quantization is too coarse, the capability of the fronthal link is not fully utilized, the overall performance is limited by the coarse quantization. Hence, an optimal tradeoff between the compression rates and the achievable sum rate must be found. The Information Bottleneck (IB) method is an effective method to find this tradeoff as well as the corresponding optimized quantizer.

Consider three variables $X \to Y \to \hat{Y}$ forming a Markov chain, where $\hat{Y}$ is the compression of Y. When we want the variable $\hat{Y}$ to compress Y as much as possible (smaller $I(Y;\hat{Y})$), while $\hat{Y}$ captures as much of the information about X as possible (larger $I(X;\hat{Y})$), the IB method is an useful tool. According to [8] and [9], with the IB method we can compute the maximized $I(X;\hat{Y})$ as the function of the compression rate $I(Y;\hat{Y})$. In detail, the function

$$I(c) = \sup_{I(Y;\hat{Y}) \le c} I(X;\hat{Y})$$

can be computed and plotted. It has been proved to be a concave and increasing function for $c \in [0, H(\hat{Y})]$. The IB method is a *deterministic annealing* approach such that the whole curve $I(c)$ is obtained through a third parameter $\beta,\ \beta > 0$, where $1/\beta = \frac{\mathrm{d}I(c)}{\mathrm{d}c}$ corresponds to the slope of the curve at the point $(c, I(c))$. Actually β is the *Lagrange Multiplier* used in the IB method. We call it the tradeoff factor between the compression rate c and the objective mutual information. By choosing an arbitrary $\beta > 0$ as the input of the IB method, the point on the tradeoff curve with slope $1/\beta$ can be outputted. Since $I(c)$ is concave and increasing, the output compression rate c of the IB method increases with the input β, the whole tradeoff curve can be obtained by ranging the value of β from 0 to infinity and running the IB method repeatedly. After obtaining this tradeoff curve, we can use the Bi-Section method to find the specific value of β such that at this point the compression rate c can be supported and the objective mutual information is maximized.

In the uplink of C-RAN, a joint optimization among all the quantizers has to be performed. In this section, we assume the we extend the IB method to a so-called Alternating Information Bottleneck (AIB) method, in order to find the optimal tradeoff between all the compression rates and the sum rate of C-RAN. Then we propose an Alternating Bi-Section method, based on the AIB method, to locate the optimal point where the optimal quantizers' design can be found.

For the ease of illustration, we start with the 2 MSs and 2 RUs case to show the optimization scheme and its convergence. At the end of this section, we will show our proposed optimization algorithms can be easily extended to the case with more devices. According to (1), the problem becomes

$$\begin{aligned} \max_{P_{\hat{Y}_1|Y_1} P_{\hat{Y}_2|Y_2}} \quad & w_1 I(X_1;\hat{Y}_1,\hat{Y}_2) + w_2 I(X_2;\hat{Y}_1,\hat{Y}_2|X_1), \\ \text{Subject to } & I(Y_1;\hat{Y}_1) \le C_1, \\ & I(Y_2;\hat{Y}_2|\hat{Y}_1) \le C_2. \end{aligned} \tag{4}$$

Let $R_{\mathrm{wsum}} = w_1 I(X_1;\hat{Y}_1,\hat{Y}_2) + w_2 I(X_2;\hat{Y}_1,\hat{Y}_2|X_1)$. Firstly we set up the tradeoff between the compression rate pair of 2 RUs and the corresponding maximized weighted sum rate R_{wsum} in Sec. III-A. Then in Sec. III-B we locate the specific point where the constraints in (4) is satisfied and the weighted sum rate is maximized.

A. The Alternating Information Bottleneck (AIB) method – Setting up the tradeoff between the compression rate pair $\Big(I(Y_1;\hat{Y}_1), I(Y_2;\hat{Y}_2|\hat{Y}_1)\Big)$ and the maximized weighted sum rate R_{wsum}, through the tradeoff factor pair (β_1,β_2)

Since the two quantizers need to be optimized jointly, the IB method cannot be used directly to numerically compute the optimal tradeoff. However we note that when one quantizer is fixed, the remaining part just becomes into the form such that the IB method can be readily used.

1. When the first quantizer $P_{\hat{Y}_1|Y_1}$ is fixed, then we need to find the optimal tradeoff between the compression rate $c_2 = I(Y_2;\hat{Y}_2|\hat{Y}_1)$ and $\max\limits_{P_{\hat{Y}_2|Y_2}} R_{\mathrm{wsum}}$. Because of the chain rule,

$$\begin{aligned} R_{\mathrm{wsum}} =& w_1 I(X_1;\hat{Y}_1,\hat{Y}_2) + w_2 I(X_2;\hat{Y}_1,\hat{Y}_2|X_1) \\ =& w_1 I(X_1;\hat{Y}_1) + w_2 I(X_2;\hat{Y}_1|X_1) \\ &+ w_1 I(X_1;\hat{Y}_2|\hat{Y}_1) + w_2 I(X_2;\hat{Y}_2|\hat{Y}_1,X_1). \end{aligned} \tag{5}$$

Thus, when the quantizer $P_{\hat{Y}_1|Y_1}$ is fixed, it is sufficient to compute the tradeoff between $I(Y_2;\hat{Y}_2|\hat{Y}_1)$ and $\max\limits_{P_{\hat{Y}_2|Y_2}} \Big(w_1 I(X_1;\hat{Y}_2|\hat{Y}_1) + w_2 I(X_2;\hat{Y}_2|\hat{Y}_1,X_1)\Big)$. Then it is reduced to the similar problem that has been solved in [9] by the IB method. This algorithm *Function IB2* will be shown in the full version paper. In this function, the fixed quantizer $P^{\text{fixed}}_{\hat{Y}_1|Y_1}$ is the input, which is an local invariable when optimizing the quantizer $P_{\hat{Y}_2|Y_2}$. The *Lagrange Multiplier* $\beta_2 > 0$ is the tradeoff factor. We obtain different tradeoff points $\left\{ I(Y_2;\hat{Y}_2|\hat{Y}_1),\ \max\limits_{P_{\hat{Y}_2|Y_2}} R_{\mathrm{wsum}} \right\}$ by inserting different values of β_2 to *Function IB2*. In Fig. 2, we fix an valid $P_{\hat{Y}_1|Y_1}$, by ranging β_2 from 0.1 to 50 and running *function IB2* repeatedly, the concave tradeoff curve in blue can be obtained.

2. Similarly, when the second quantizer $P_{\hat{Y}_2|Y_2}$ is fixed, the tradeoff points $\left\{ I(Y_1;\hat{Y}_1),\ \max\limits_{P_{\hat{Y}_1|Y_1}} R_{\mathrm{wsum}} \right\}$ can also be obtained with the IB method, as summarized in *Function IB1* (will be available in the full version). The fixed quantizer $P^{\text{fixed}}_{\hat{Y}_2|Y_2}$ is the input and $\beta_1 > 0$ is the tradeoff factor. The concave tradeoff curve is plotted in red in Fig. 2.

Then we go back to the original problem (4). Since the two quantizers should be optimized jointly, the optimization of one quantizer always depends on the optimization results of the other. Hence, we can obtain the tradeoff between the compression rate pair $\Big(I(\hat{Y}_1;Y_1),\ I(\hat{Y}_2;Y_2|\hat{Y}_1)\Big)$ and $\max\limits_{P_{Y_1\hat{|}Y_1} P_{Y_2\hat{|}Y_2}} R_{\mathrm{wsum}}$ by running these two functions alternatively, such that the optimized quantizer outputted from one IB function is the input fixed quantizer of the other, until reaching the convergence. This Alternating Information Bottleneck (AIB) method in summarized in *Function AIB* (will be available in the full version).

The AIB method will definitely converge to the local optimal point. Since the IB function $i,\ i \in \{1,2\}$, will converge to the point where R_{wsum} is maximized, for the current fixed $P_{\hat{Y}_j|Y_j}$. Then in the AIB method, we set the

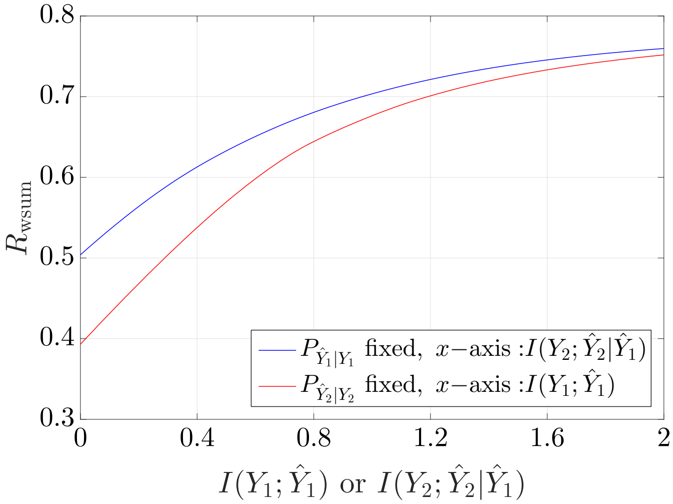

Fig. 2. Tradeoff between the sum rate and the compression rates. BPSK modulation, $w_1 = w_2 = 1$, $h_{11} = 1$, $h_{12} = 0.4$, $h_{21} = 0.6$, $h_{22} = 0.9$, $P_1 = 1$, $P_2 = 0.5$, $\sigma_n^2 = 1$, $|\hat{\mathcal{Y}}_1| = |\hat{\mathcal{Y}}_2| = 8$, $\epsilon_1 = 0.0003$, $\beta_1,\ \beta_2 \in [0.1, 50]$.

fixed $P_{\hat{Y}_j|Y_j}$ and the optimized $P_{\hat{Y}_i|Y_i}$ as the starting point of the IB function j for the optimization of $P_{\hat{Y}_j|Y_j}$, in order to further maximize R_{wsum} Hence, for specific compression rate pair, the corresponding R_{wsum} will not be decreased in each iteration and converge to the local optimal point. Since the problem is generally non-convex, similar to the IB method, we can try different initial point in the AIB method to get better results. When the AIB method converges, we can obtain the specific tradeoff point between the compression rate pair $\Big(I(Y_1;\hat{Y}_1),\ I(Y_2;\hat{Y}_2|\hat{Y}_1)\Big)$ associated with the input tradeoff factor pair (β_1,β_2), and the corresponding maximized weighted sum rate R_{wsum}, as well as the optimized quantizers. By choosing different values of β_1 and β_2 and running the AIB method repeatedly, different tradeoff points can be obtained.

B. The Alternating Bi-Section method – Locating the optimal tradeoff point

After setting up the tradeoff numerically through the tradeoff pair (β_1,β_2), we need to locate the point such that the constraints in (4) are simultaneously fulfilled. Obviously, in order to fully utilize the fronthaul links, we need to find the point where $I(Y_1;\hat{Y}_1) = C_1$ and $I(Y_2;\hat{Y}_2|\hat{Y}_1) = C_2$, then its corresponding maximized weighted sum rate R_{wsum} and the quantizers $P_{\hat{Y}_1|Y_1}$, $P_{\hat{Y}_2|Y_2}$ are the solution for (4). We can find this point by exhaustively inserting different tradeoff factor pairs (β_1,β_2) until finding the point where $c_1 = C_1$ and $c_2 = C_2$. Obviously this approach is rather inefficient. In the scenario of one quantizer, since the compression rate c increases with the input value of β, we can use the Bi-Section method to find the specific value of β such that at the point of the tradeoff curve with slope $1/\beta$, c equals to the capacity of the constraint link. The details can be found in [9]. While in C-RAN, there're multiple quantizers, the resultant compression rate of a quantizer also depends on the compression rates of other quantizers, the Bi-Section method can not be directly used to locate each tradeoff factor individually. In Fig. 3, we run the AIB function with different input tradeoff factor

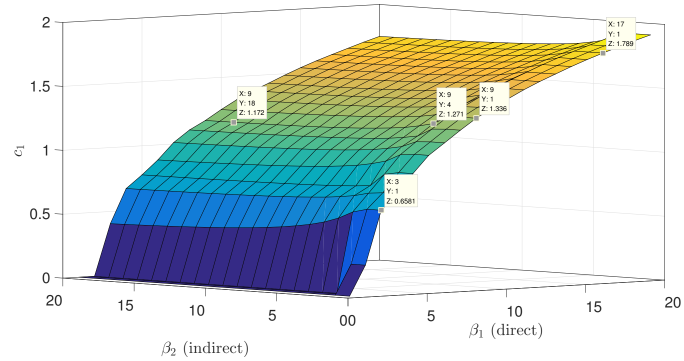

Fig. 3. Relationship between the input tradeoff factor pair (β_1,β_2) with the output compression rate c_1. The same channel setup of Fig. 2 is assumed.

pairs (β_1,β_2), and plot the output compression rate c_1 as the function of it. We see that c_1 depends mainly on the value of β_1, such that when β_2 is fixed, c_1 increases with the value of β_1, which is the same as the conventional one quantizer case. We say c_1 is directly associated with β_1. However, since the design of the quantizers affects each other, β_2 also slightly influences c_1 in a non-linear way, see the marked points in figure. We say c_1 is indirectly associated with β_2. When we use the Bi-Section method to locate β_1 and β_2 individually, such that we locate β_1 where $c_1(\beta_1) = C_1$, then we fix this β_1 and locate β_2 until reaching $c_2(\beta_2) = C_2$, the newly located β_2 (c_1's indirect tradeoff factor) will make c_1 slightly deviate from the previous value and vice versa. Hence, the tradeoff factor pair should be located jointly instead of individually with the Bi-Section method. Similar to the AIB method, we propose an Alternating approach to achieve this goal, called Alternating Bi-Section method. It incorporates the AIB method. For a specific target compression rate pair (C_1, C_2), it can compute the corresponding associated tradeoff factor pair (β_1,β_2). This Alternating Bi-Section method will be available in the full version of the paper.

C. Extension to more MSs and RUs with multiple antennas

...similar to the 2 MSs case, will be available in the full version.

IV. Optimization Scheme for the Capacity Allocation

We assume that the capacity allocated to each fronthaul link is fixed in the previous section, and proposed the AIB method and the Alternating Bi-Section method for joint optimization of the quantizers and location of the optimal tradeoff point. In this section, we assume the fronthaul links are constrained by the sum capacity, and address the problem of optimizing resource allocation. We propose the algorithm by combining the AIB method, the Alternating Bi-Section method with the Outer Linearization Method (OLM). The algorithm is sketched as below.

1. Start with a random valid capacity allocation, $\boldsymbol{C}^{(0)} = \Big(C_1^{(0)}, C_2^{(0)}, ..., C_L^{(0)}\Big)$, such that $\sum_{i=1}^{L} C_i^{(0)} = C_{\text{sum}}$. Set $k = 0$, $f_{\text{LB}} = -1$ and f_{UB} to be large enough. Set δ be the desired tolerance.

At iteration k, repeat step 2 to step 4 until $f_{\text{UB}} - f_{\text{LB}} \leq \delta$.

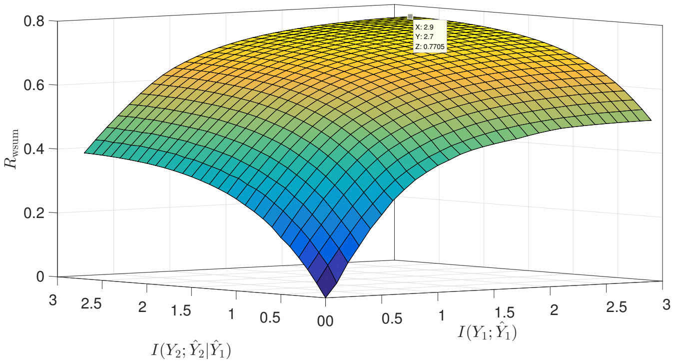

Fig. 4. The tradeoff plane. $h_{11} = 1,\ h_{12} = 0.4,\ h_{21} = 0.6,\ h_{22} = 0.9,\ P_1 = 1,\ P_2 = 0.5,\ \sigma_n^2 = 1,\ |\hat{\mathcal{Y}}_1| = |\hat{\mathcal{Y}}_2| = 8,\ w_1 = w_2 = 1.$

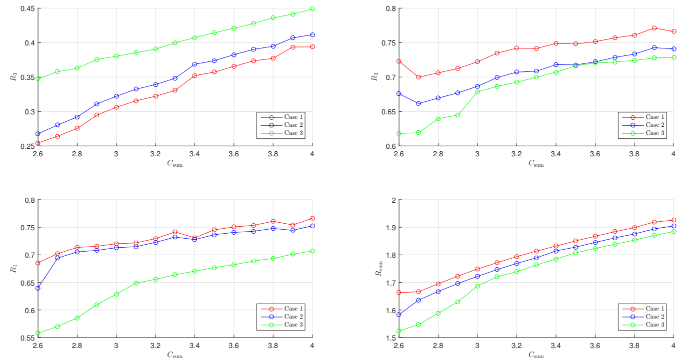

Fig. 5. Relationship between achievable rates with sum capacity of fronthaul. $h_{11} = 1,\ h_{12} = 0.3,\ h_{13} = 0.2,\ h_{21} = 0.2,\ h_{22} = 1,\ h_{23} = 0.3,\ h_{31} = 0.2,\ h_{32} = 0.1,\ h_{33} = 0.5,\ \sigma_n^2 = 1,\ |\hat{\mathcal{Y}}_1| = |\hat{\mathcal{Y}}_2| = |\hat{\mathcal{Y}}_3| = 8.$

2. Use the Alternating Bi-Section method to compute the tradeoff factor vector $\boldsymbol{\beta}^{(k)} = \left(\beta_1^{(k)}, \beta_2^{(k)}, ..., \beta_L^{(k)}\right)$ associated with the current capacity allocation $\boldsymbol{C}^{(k)}$.

3. Insert $\boldsymbol{\beta}^{(k)}$ to the AIB method, then compute current maximized weighted sum rate $R_{\mathrm{wsum}}^{(k)}$. Set $f_{\mathrm{LB}} = R_{\mathrm{wsum}}^{(k)}$ and the subgradient $\boldsymbol{g}^{(k)} = \left(1/\beta_1^{(k)}, 1/\beta_2^{(k)}, ..., 1/\beta_L^{(k)}\right)$ and $b^{(k)} = R_{\mathrm{wsum}}^{(k)} - \boldsymbol{C}^{(k)} \cdot (\boldsymbol{g}^{(k)})^T$.

4. Solve the linear problem

$$\max_{s,\boldsymbol{C}} s$$
$$s.t.\ \boldsymbol{C} \cdot (\boldsymbol{g}^{(\ell)})^T + b^{(\ell)} \geq s,\ \ \ell = 0, 1, ..., k-1,$$
$$\sum_{i=1}^{L} C_i = C_{\mathrm{sum}}$$

Let $(s^*, \boldsymbol{C}^*)$ be the maximizer, set $f_{\mathrm{UB}} = s^*$ and $\boldsymbol{C}^{(k+1)} = \boldsymbol{C}^*$. Set $k = k + 1$.

The validation and explanation of the algorithm will be given in the full version.

V. Simulation Results

In this section, we evaluate the performance of the proposed algorithms and use them to study C-RAN.

For the 2 MSs and 2 RUs case, we set different target compression rate pairs (c_1, c_2), and use the Alternating Bi-Section method to obtain the tradeoff plane between the compression rate pair $\left(I(Y_1;\hat{Y}_1), I(Y_2;\hat{Y}_2|\hat{Y}_1)\right)$ and the corresponding maximized weighted sum rate R_{wsum}, as shown in Fig 4. It is a convex and increasing plane of the compression rate pair. The maximum achievable sum rate increases when either compression rate increases. Thus if the capacities of the fronthaul links become larger, the sum rate will be more and more close to the theoretical limit 0.7892.

Then we consider a 3MSs-3RUs C-RAN. We set $w_3 = 3,\ w_1 = w_2 = 1$. At first we use the proposed algorithms to optimize the quantizers as well as the capacity allocation, in order to maximize the sum rate (case 1). Then we fix this capacity allocation, and use the AIB method and the Alternating Bi-Section method to optimize the quantizers only, so as to maximize the weighted sum rate, under the current capacity allocation (case 2). At last, we optimize both the quantizers and the capacity allocation for maximizing the weighted sum rate (case 3). The results is shown in Fig. 5. From the figure we see that when the quantizers and capacity allocation are optimized in order to maximize the sum rate, the individual rate of the third user R_3 is the smallest, while the achievable sum rate is maximized. Moreover, R_1 and R_2 are the largest in these 3 cases. While when we want to put more weight on R_3 by setting $w_3 = 3$, only optimizing the quantizers is not sufficient, the improvement of R_3 in case 2 compared to the former case is not significant. This is because the received signals at different RUs are the superposition of the signals from all users, only optimizing the quantization will not impose a prominent impact on individual rates. In order to further improve the individual rate with larger weight, it is necessary to consider a simultaneous optimization of both capacity allocation and compression. From the figure, we see that by comparing with case 1, the improvement of R_3 in case 3 is much more significant than that of case 2. While this improvement is at the cost of the larger decrease of R_1, R_2 and sum rate R_{sum}.

We consider the same model of Fig. 5, and assume the sum capacity available is 3 bits/cu, the optimal capacity allocation obtained from the proposed algorithm for different optimization objectives is shown in Fig. 6. We see that when the quantizers and the capacity allocation are optimized for maximizing the sum rate, only 18% of the capacity is allocated to the third RU. While when we want to maximize the weighted sum rate ($w_3 = 3,\ w_1 = w_2 = 1$), 38% of the capacity should be allocated to the third RU. The reason is that the signal from the third user at the third RU is the strongest, while at the first RU it is the weakest. Moreover, the observation of the superposed signal is more reliable at the first and second RU than that at the third RU. Thus, when the capacity allocation is optimized for maximizing the sum rate, the capacity allocated to the third RU should be the least, while it should be the most when the achievable rate of the third user has larger weight.

More numerical results will be available in the full version of the paper.

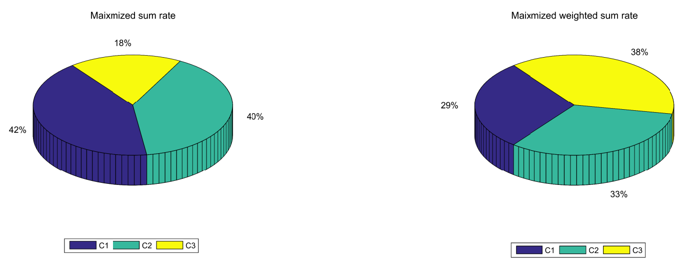

Fig. 6. Optimal capacity allocation for different objectives

VI. Conclusion

In this paper we proposed the Alternating Information Bottleneck (AIB) method, which extends the conventional Information Bottleneck (IB) method to the multi-quantizer case. It can numerically compute the optimal tradeoff between the multiple compression rates and the objective mutual information. Then we proposed the Alternating Bi-Section method, which incorporates the AIB method. It is an effective tool to design optimal quantizers when the multiple compression rates are constrained by limited capacities. Based on these two algorithms, the optimization for the capacity allocation is proposed. The algorithms are suitable for the centralized optimization of the compression step in C-RAN.

Reference

[1] China Mobile. C-ran: The road towards green ran, white paper, ver. 2.5. *China Mobile Research Institute*, Oct 2011.

[2] Seok-Hwan Park, Osvaldo Simeone, Onur Sahin, and Shlomo Shamai (Shitz). Fronthaul compression for cloud radio access networks. *IEEE Signal Processing Magazine*, Nov 2014.

[3] T. Cover and A. E. Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory*, 25(5), Sep 1979.

[4] Sung Hoon Lim, Young-Han Kim, Abbas El Gamal, and Sae-Young Chung. Noisy network coding. *IEEE Trans. Inf. Theory*, 57(5), May 2011.

[5] A. D. Wyner and Jacob Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory*, 22(1), Jan 1976.

[6] Yuhan Zhou and Wei Yu. Optimized backhaul compression for uplink cloud radio access network. *IEEE Journal on Selected Areas in Communications*, 32(6), Jun 2014.

[7] Seok-Hwan Park, Osvaldo Simeone, Onur Sahin, and Shlomo Shamai (Shitz). Joint decompression and decoding for cloud radio access networks. *IEEE Signal Processing Letters*, 20(5), May 2013.

[8] Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. *The 37th annual Allerton Conference on Communication, Control, and Computing*, Sep 1999.

[9] Georg Christoph Zeitler. Dissertation: Low-precision quantizer design for communication problems. *http://mediatum.ub.tum.de/doc/1092069/1092069.pdf*, Nov 2011.

[10] Andreas Winkelbauer. Dissertation: Blind performance estimation and quantizer design with applications to relay networks. Dec 2014.

[11] M. S. Bazaara, H. D. Sherali, and C. M. Shetty. Nonlinear programming. *Hoboken: John Wiley Sons, Inc.*, 2006.

[12] H. Boche and M. Wiczanowski. Stability optimal transmission policy for the multiple antenna multiple access channel in the geometric view. *EURASIP Signal Processing Journal*, Aug 2006.

[13] Shlomo Shamai (Shitz). On cloud radio access networks: Information theoretic considerations. *Plenary Talk of ISWCS 2014*, Aug 2014.