Optimizing Over-the-Air Computation in IRS-Aided C-RAN Systems

Daesung Yu, Seok-Hwan Park
Dept. of Elect. Engineering
Jeonbuk National University
Jeonju, Korea
{imcreative93, seokhwan}@jbnu.ac.kr

Osvaldo Simeone
KCLIP Lab, Centre for Telecomm. Research
Dept. of Engineering
King's College London
London, UK
osvaldo.simeone@kcl.ac.uk

Shlomo Shamai (Shitz)
Dept. of Elect. Engineering
Technion
Haifa, Israel
sshlomo@ee.technion.ac.il

Abstract—**Over-the-air computation (AirComp) is an efficient solution to enable federated learning on wireless channels. Air-Comp assumes that the wireless channels from different devices can be controlled, e.g., via transmitter-side phase compensation, in order to ensure coherent on-air combining. Intelligent reflecting surfaces (IRSs) can provide an alternative, or additional, means of controlling channel propagation conditions. This work studies the advantages of deploying IRSs for AirComp systems in a large-scale cloud radio access network (C-RAN). In this system, worker devices upload locally updated models to a parameter server (PS) through distributed access points (APs) that communicate with the PS on finite-capacity fronthaul links. The problem of jointly optimizing the IRSs' reflecting phases and a linear detector at the PS is tackled with the goal of minimizing the mean squared error (MSE) of a parameter estimated at the PS. Numerical results validate the advantages of deploying IRSs with optimized phases for AirComp in C-RAN systems.**

Index Terms—**Over-the-air computation, C-RAN, intelligent reflecting surface.**

I. INTRODUCTION

Federated learning is an emerging distributed learning paradigm in which mobile devices collaboratively train a machine learning model while preserving the privacy of local data sets [1]. In the presence of latency and bandwidth constraints, the implementation of federated learning on wireless systems is challenging if many workers, or devices, are involved. A potential solution to this problem is over-the-air computation (AirComp), which leverages the superposition property of the multiple access channel (MAC) from worker devices to a parameter server (PS) to allow for simultaneous transmissions from multiple devices [2]–[4]. It was reported in [3] that Air-Comp outperforms a conventional multiple access technique in terms of test accuracy, and that the gain is particularly significant at low transmit power and large number of workers.

AirComp assumes that the wireless channels from different devices can be controlled, e.g., via transmitter-side phase compensation, in order to ensure coherent on-air combining
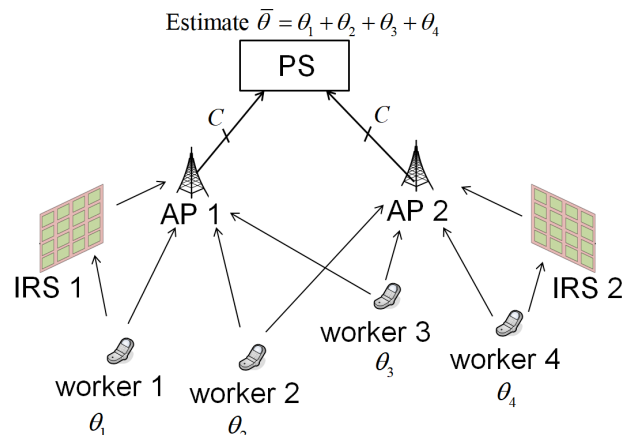
Fig. 1. Over-the-air computation system in an IRS-aided C-RAN with $N_W = 4$, $N_A = 2$ and $N_I = 2$.

[5]. To alleviate this problem, the work [6] considered a deployment of intelligent reflecting surfaces (IRSs). IRSs, also referred to as reconfigurable intelligent surfaces, can be controlled through integrated electronics in order to shape their response to impinging electromagnetic waves [7]. This enables the modification of the propagation channel between nearby transceivers. As a result, IRSs are considered as a cost-effective solution to improve spectral and energy efficiency of wireless systems [8]–[11]. As examples of recent works on IRSs, references [9] and [10] addressed the joint design of downlink beamforming and IRSs' phases for interference management in multi-user [9] and multi-cell systems [10]. Reference [8] analyzed the number of reflecting elements of IRSs needed to beat conventional wireless relaying techniques (see also [12]). Finally, an information-theoretic study was provided in [13].

In this work, we study the advantages of deploying IRSs for AirComp systems. Unlike [6], which focused on a MAC channel where workers directly communicate with a PS, we consider the large-scale cloud radio access network (C-RAN) illustrated in Fig. 1, in which the workers upload local models to the PS through distributed access points (APs). The APs, or remote radio heads (RRHs), in C-RAN send the received signals to the PS on fronthaul links. The fronthaul

links have finite capacity, requiring fronthaul quantization and compression [14]. We tackle the problem of jointly optimizing the IRSs' reflecting phases and a linear detector at the PS with the goal of minimizing the mean squared error (MSE) of a parameter estimated at the PS. Due to the non-convexity of the problem, we propose an iterative algorithm that alternately updates the IRSs' phases and the linear detector. Via numerical results, we validate the advantages of deploying IRSs with optimized phases for AirComp in C-RAN systems.

II. SYSTEM MODEL

As illustrated in Fig. 1, we consider an over-the-air computation task performed on a C-RAN system. In the system, N_W single-antenna worker devices send locally updated models to a PS through N_A single-antenna APs. Each AP is connected to the PS via a fronthaul link, which we model as a digital link of capacity C bit/sample [14]. We define the sets $\mathcal{N}_W = \{1, 2, \ldots, N_W\}$ and $\mathcal{N}_A = \{1, 2, \ldots, N_A\}$ for the workers' and APs' indices, respectively.

A. Over-the-Air Computation Model

We focus on the transmission at a specific time slot where each worker $k \in \mathcal{N}_W$ sends a scalar parameter θ_k, and the PS estimates a function $f(\boldsymbol{\theta})$ of the transmitted parameters $\boldsymbol{\theta} = \{\theta_k\}_{k \in \mathcal{N}_W}$. The parameter θ_k can be an element of the gradient vector [3] or the local model [4] updated at worker k using its local dataset. The PS typically estimates the weighted sum $f(\boldsymbol{\theta}) = \sum_{k \in \mathcal{N}_W} w_k \theta_k$, with $w_k = S_k / (\sum_{l \in \mathcal{N}_W} S_l)$, where S_k denotes the number of training samples at device k [4]. To simplify the discussion, we assume $S_k = S$ for all $k \in \mathcal{N}_W$, and that the target parameter denoted by $\bar{\theta}$ is given by the sum

$$\bar{\theta} = f(\boldsymbol{\theta}) = \sum_{k \in \mathcal{N}_W} \theta_k. \tag{1}$$

We also assume that the parameters θ_k are independent, and we define the power of parameter θ_k as $\mathrm{E}[|\theta_k|^2] = \sigma_{\theta,k}^2$. Thus, the target parameter $\bar{\theta}$ has power $\mathrm{E}[|\bar{\theta}|^2] = \sum_{k \in \mathcal{N}_W} \sigma_{\theta,k}^2$.

B. Channel Model

To assist edge communication from the workers to the APs, we assume the presence of N_I IRSs [6] in the network. Each IRS has n_I reflecting elements, whose reflecting phases are dynamically adjusted to adapt to the instantaneous channel state information (CSI). We define the set $\mathcal{N}_I = \{1, 2, \ldots, N_I\}$ for the IRSs' indices.

Under a flat-fading channel model, the received signal y_i of AP i can be written as

$$y_i = \sum_{k \in \mathcal{N}_W} h_{i,k} x_k + z_i, \tag{2}$$

where x_k is the signal transmitted by worker k; $h_{i,k}$ denotes the channel coefficient from worker k to AP i; and $z_i \sim \mathcal{CN}(0, \sigma_z^2)$ represents the additive noise. The signal x_k satisfies the transmit power constraint $\mathrm{E}[|x_k|^2] \leq P$.

Due to the presence of IRSs, the channel coefficient $h_{i,k}$ is modelled as [8]–[10]

$$h_{i,k} = \sqrt{\rho_{d,i,k}} h_{d,i,k} + \sum_{j \in \mathcal{N}_I} \sqrt{\rho_{r,i,j,k}} \mathbf{g}_{i,j}^H \boldsymbol{\Theta}_j \mathbf{h}_{r,j,k}, \tag{3}$$

where $h_{d,i,k}$ denotes the small-scale fading channel from worker k to AP i; $\mathbf{g}_{i,j} \in \mathbb{C}^{n_I \times 1}$ represents the small-scale fading channel vector from IRS j to AP i; $\mathbf{h}_{r,j,k} \in \mathbb{C}^{n_I \times 1}$ is the small-scale fading channel vector from worker k to IRS j; $\rho_{d,i,k}$ denotes the path-loss of the direct link from worker k to AP i; $\rho_{r,i,j,k}$ is the path-loss of the composite link from worker k to AP i through IRS j; and $\boldsymbol{\Theta}_j$ is a diagonal matrix that represents the reflecting operation of IRS j, which is defined as

$$\boldsymbol{\Theta}_j = \mathrm{diag}\left(\{e^{j\phi_{j,m}}\}_{m=1}^{n_I}\right), \tag{4}$$

where $\phi_{j,m} \in [0, 2\pi)$ denotes the reflecting phase of the mth element of IRS j.

We model the path-loss $\rho_{d,i,k}$ between worker k and AP i as $\rho_{d,i,k} = c_0 \cdot \mathrm{D}(\mathbf{p}_{W,k}, \mathbf{p}_{A,i})^{-\eta}$, where $\mathrm{D}(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|_2$ is the Euclidean distance in meter between the two input vectors, $\mathbf{p}_{W,k}$ and $\mathbf{p}_{A,i}$ denote the position vectors of worker k and AP i, respectively, η is the path-loss exponent, and c_0 denotes the path-loss at the reference distance of 1 m. For the path-loss $\rho_{r,i,j,k}$ of the composite channel from worker k to AP i through IRS j, we adopt the sum-distance model [7] which models $\rho_{r,i,j,k}$ as

$$\rho_{r,i,j,k} = c_0 \left(\mathrm{D}(\mathbf{p}_{W,k}, \mathbf{p}_{I,j}) + \mathrm{D}(\mathbf{p}_{I,j}, \mathbf{p}_{A,i})\right)^{-\eta}, \tag{5}$$

where $\mathbf{p}_{I,j}$ denotes the position vector of IRS j.

III. OVER-THE-AIR COMPUTATION IN IRS-AIDED C-RAN

In this section, we illustrate the operations at the worker devices, the APs, and the PS in the IRS-aided C-RAN system described in Sec. II.

A. Transmission at Worker Devices

Without claim of optimality (see [15]), we assume that each worker k uses the maximum transmit power P, so that the transmit signal x_k is given as

$$x_k = \alpha_k \theta_k, \tag{6}$$

with the coefficient $\alpha_k = (P/\sigma_{\theta,k}^2)^{1/2}$. We note that this does not require CSI at worker devices.

B. Quantization at APs

AP i sends a quantized version of the received signal y_i to the PS through a fronthaul link of capacity C bit/sample. Under the assumptions that the updated model vectors have a sufficiently large dimension, the quantized signal denoted by $\hat{y}_i$ can be modelled as [14], [16]

$$\hat{y}_i = y_i + q_i, \tag{7}$$

where q_i models the quantization distortion as being independent of y_i and distributed as $q_i \sim \mathcal{CN}(0, \omega_i)$. According to

standard rate-distortion theoretic results [17], the quantization noise power ω_i satisfies the condition

$$I(y_i; \hat{y}_i) = \log_2\left(1 + \frac{\sigma_{y,i}^2}{\omega_i}\right) \leq C, \tag{8}$$

where $\sigma_{y,i}^2$ denotes the variance of the received signal y_i given as

$$\sigma_{y,i}^2 = \sum_{k \in \mathcal{N}_W} |h_{i,k}|^2 P + \sigma_z^2. \tag{9}$$

The minimum distortion power ω_i that satisfies the condition (8) is given as

$$\omega_i = \sigma_{y,i}^2/(2^C - 1). \tag{10}$$

Note that the optimal distortion level (10) is a function of the reflecting phases $\phi = \{\phi_{j,m}\}_{j \in \mathcal{N}_I, m \in \{1,\ldots,n_I\}}$, since ϕ affects the channel coefficients $h_{i,k}$ as seen in (3).

C. Estimation at PS

Based on the received quantized signals $\{\hat{y}_i\}_{i \in \mathcal{N}_A}$, the PS estimates the target parameter $\bar{\theta}$ in (1). To elaborate, let us define a vector $\hat{\mathbf{y}} = [\hat{y}_1 \hat{y}_2 \cdots \hat{y}_{N_A}]^T$ which stacks the quantized signals. Then, the vector $\hat{\mathbf{y}}$ can be expressed as

$$\hat{\mathbf{y}} = \sum_{k \in \mathcal{N}_W} \alpha_k \mathbf{h}_k \theta_k + \mathbf{z} + \mathbf{q}, \tag{11}$$

where we have defined the vectors $\mathbf{h}_k = [h_{1,k} h_{2,k} \cdots h_{N_A,k}]^T$, $\mathbf{z} = [z_1 z_2 \cdots z_{N_A}]^T \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2 \mathbf{I})$ and $\mathbf{q} = [q_1 q_2 \cdots q_{N_A}]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega})$ with $\mathbf{\Omega} = \text{diag}(\{\omega_i\}_{i \in \mathcal{N}_A})$.

The channel vector $\mathbf{h}_k \in \mathbb{C}^{N_A \times 1}$ from worker k to all the APs can be written as a function of the IRSs' phases ϕ as

$$\mathbf{h}_k = \mathbf{h}_{d,k} + \sum_{j \in \mathcal{N}_I} \mathbf{R}_{r,j,k} \mathbf{G}_j \text{diag}(\mathbf{h}_{r,j,k})\mathbf{v}_j, \tag{12}$$

where the matrices $\mathbf{R}_{r,j,k} \in \mathbb{C}^{N_A \times N_A}$, $\mathbf{G}_j \in \mathbb{C}^{N_A \times n_I}$, and the vectors $\mathbf{h}_{d,k} \in \mathbb{C}^{N_A \times 1}$, $\mathbf{v}_j \in \mathbb{C}^{n_I \times 1}$ are defined as $\mathbf{R}_{r,j,k} = \text{diag}(\{\rho_{r,i,j,k}^{1/2}\}_{i \in \mathcal{N}_A})$, $\mathbf{G}_j = [\mathbf{g}_{1,j} \cdots \mathbf{g}_{N_A,j}]^H$, $\mathbf{h}_{d,k} = [\rho_{d,1,k}^{1/2} h_{d,1,k} \cdots \rho_{d,N_A,k}^{1/2} h_{d,N_A,k}]^T$, and $\mathbf{v}_j = [e^{j\phi_{j,1}} \cdots e^{j\phi_{j,n_I}}]^T$, respectively. Note that the optimization of the phases $\{\phi_{j,m}\}_{m=1}^{n_I}$ of IRS j is equivalent to that of the vector $\mathbf{v}_j$ as long as the conditions

$$|\mathbf{v}_j(m)|^2 = 1 \tag{13}$$

are satisfied for all $m \in \{1, \ldots, n_I\}$, where $\mathbf{v}_j(m)$ denotes the mth element of $\mathbf{v}_j$. From the vector $\mathbf{v}_j$, each phase $\phi_{j,m}$ can be obtained as $-\angle\mathbf{v}_j(m)$.

We assume that the PS performs a linear estimation of the target parameter $\bar{\theta}$ from $\hat{\mathbf{y}}$. Accordingly, an estimate $\hat{\bar{\theta}}$ of $\bar{\theta}$ is given as

$$\hat{\bar{\theta}} = \mathbf{f}^H \hat{\mathbf{y}}, \tag{14}$$

with a linear detection vector $\mathbf{f} \in \mathbb{C}^{N_A \times 1}$.

For given phases ϕ, i.e., $\mathbf{v} = \{\mathbf{v}_j\}_{j \in \mathcal{N}_I}$, and linear detection vector $\mathbf{f}$, the MSE between the estimate $\hat{\bar{\theta}}$ and the target parameter $\bar{\theta}$ is evaluated as

$$\mathbf{e}(\mathbf{v}, \mathbf{f}) = \mathbb{E}\left[|\hat{\bar{\theta}} - \bar{\theta}|^2\right] \tag{15}$$

$$= \sum_{k \in \mathcal{N}_W} |\alpha_k \mathbf{f}^H \mathbf{h}_k - 1|^2 \sigma_{\theta,k}^2 + \mathbf{f}^H\left(\sigma_z^2 \mathbf{I} + \mathbf{\Omega}\right)\mathbf{f}.$$

IV. OPTIMIZATION

We tackle the problem of jointly optimizing the IRSs' reflecting phases $\mathbf{v}$ and the linear detection vector $\mathbf{f}$ of the PS with the goal of minimizing the MSE $\mathbf{e}(\mathbf{v}, \mathbf{f})$ in (15) while satisfying the unit modulus constraints (13). The problem can be stated as

$$\underset{\mathbf{v}, \mathbf{f}}{\text{minimize}}\ \ \mathbf{e}(\mathbf{v}, \mathbf{f}) \tag{16a}$$

$$\text{s.t.}\ \ |\mathbf{v}_j(m)|^2 = 1,\ j \in \mathcal{N}_I,\ m \in \{1, \ldots, n_I\}. \tag{16b}$$

Since it is difficult to jointly optimize the variables $\mathbf{v}$ and $\mathbf{f}$, we propose an iterative algorithm that alternately optimizes one variable while fixing other.

If we fix the IRSs' phases $\mathbf{v}$ in problem (16), finding the optimal detector $\mathbf{f}$ becomes an unconstrained quadratic optimization problem, whose closed-form solution is given as

$$\mathbf{f} = \left(\sum_{k \in \mathcal{N}_W} P \mathbf{h}_k \mathbf{h}_k^H + \sigma_z^2 \mathbf{I} + \mathbf{\Omega}\right)^{-1} \sum_{k \in \mathcal{N}_W} \alpha_k \sigma_{\theta,k}^2 \mathbf{h}_k. \tag{17}$$

To tackle the problem of optimizing the IRSs' phases $\mathbf{v}$ for fixed $\mathbf{f}$, we remove the terms that are not dependent on the IRSs' phases from the cost function. Stating the obtained problem with respect to a stacked vector $\bar{\mathbf{v}} = [\mathbf{v}_1^H \mathbf{v}_2^H \cdots \mathbf{v}_{N_I}^H]^H \in \mathbb{C}^{\bar{n}_I \times 1}$ with $\bar{n}_I = n_I N_I$ yields

$$\underset{\bar{\mathbf{v}}}{\text{minimize}}\ \begin{pmatrix} \sum_{k \in \mathcal{N}_W}\left(|\mathbf{a}_k^H \bar{\mathbf{v}}|^2 + 2\Re\{b_k^* \mathbf{a}_k^H \bar{\mathbf{v}}\}\right)\sigma_{\theta,k}^2 + \\ \sum_{i \in \mathcal{N}_A, k \in \mathcal{N}_W}\left(|\mathbf{c}_{i,k}^H \bar{\mathbf{v}}|^2 + 2\Re\{d_{i,k}^* \mathbf{c}_{i,k}^H \bar{\mathbf{v}}\}\right) \end{pmatrix} \tag{18a}$$

$$\text{s.t.}\ \ |\bar{\mathbf{v}}(m)|^2 = 1,\ m \in \{1, \ldots, \bar{n}_I\}, \tag{18b}$$

where we have defined the notations $\mathbf{a}_k = \bar{\mathbf{H}}_{r,k}^H \mathbf{f} \alpha_k^* \in \mathbb{C}^{\bar{n}_I \times 1}$, $b_k = \alpha_k \mathbf{f}^H \mathbf{h}_{d,k} - 1$, $\bar{\mathbf{H}}_{r,k} = [\mathbf{H}_{r,1,k} \mathbf{H}_{r,2,k} \cdots \mathbf{H}_{r,N_I,k}] \in \mathbb{C}^{N_A \times \bar{n}_I}$, $\mathbf{H}_{r,j,k} = \mathbf{R}_{r,j,k} \mathbf{G}_j \text{diag}(\mathbf{h}_{r,j,k}) \in \mathbb{C}^{N_A \times n_I}$, $\mathbf{c}_{i,k} = (P/(2^C - 1))^{1/2} |\mathbf{f}(i)| \bar{\mathbf{H}}_{r,k}^H \mathbf{e}_i \in \mathbb{C}^{\bar{n}_I \times 1}$, and $d_{i,k} = (P/(2^C - 1))^{1/2} |\mathbf{f}(i)| \mathbf{e}_i^H \mathbf{h}_{d,k}$ with $\mathbf{f}(i)$ and $\mathbf{e}_i$ being the ith element of $\mathbf{f}$ and the ith column of an identity matrix of size N_A, respectively.

The problem (18) is non-convex due to the unit modulus constraints (18b). To handle this issue, we adopt the matrix lifting approach proposed in [6]. Accordingly, we tackle the problem (18) with respect to a matrix $\mathbf{V} \in \mathbb{C}^{(\bar{n}_I + 1) \times (\bar{n}_I + 1)}$ defined as

$$\mathbf{V} = \begin{bmatrix} \bar{\mathbf{v}} \\ 1 \end{bmatrix}\begin{bmatrix} \bar{\mathbf{v}}^H & 1 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{v}}\bar{\mathbf{v}}^H & \bar{\mathbf{v}} \\ \bar{\mathbf{v}}^H & 1 \end{bmatrix}. \tag{19}$$

The matrix $\mathbf{V}$ is subject to the constraints $\mathbf{V} \succeq \mathbf{0}$, $\text{rank}(\mathbf{V}) \leq 1$, and $\mathbf{V}(m,m) = 1$ for all $m \in \{1, 2, \ldots, \bar{n}_I + 1\}$. From

$\mathbf{V}$, the IRSs' phase vector $\bar{\mathbf{v}}$ can be recovered as the first $\bar{n}_I$ elements of the last column of $\mathbf{V}$.

We tackle (18) with respect to $\mathbf{V}$ by using the following equalities:

$$|\mathbf{a}_k^H \bar{\mathbf{v}}|^2 + 2\Re\{b_k^* \mathbf{a}_k^H \bar{\mathbf{v}}\} = \begin{bmatrix} \bar{\mathbf{v}}^H & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a}_k \mathbf{a}_k^H & b_k \mathbf{a}_k \\ b_k^* \mathbf{a}_k^H & 0 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{v}} \\ 1 \end{bmatrix}$$

$$= \mathrm{tr}\left(\begin{bmatrix} \mathbf{a}_k \mathbf{a}_k^H & b_k \mathbf{a}_k \\ b_k^* \mathbf{a}_k^H & 0 \end{bmatrix} \mathbf{V} \right), \tag{20}$$

$$|\mathbf{c}_{i,k}^H \bar{\mathbf{v}}|^2 + 2\Re\{d_{i,k}^* \mathbf{c}_{i,k}^H \bar{\mathbf{v}}\} = \begin{bmatrix} \bar{\mathbf{v}}^H & 1 \end{bmatrix} \begin{bmatrix} \mathbf{c}_{i,k} \mathbf{c}_{i,k}^H & d_{i,k} \mathbf{c}_{i,k} \\ d_{i,k}^* \mathbf{c}_{i,k}^H & 0 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{v}} \\ 1 \end{bmatrix}$$

$$= \mathrm{tr}\left(\begin{bmatrix} \mathbf{c}_{i,k} \mathbf{c}_{i,k}^H & d_{i,k} \mathbf{c}_{i,k} \\ d_{i,k}^* \mathbf{c}_{i,k}^H & 0 \end{bmatrix} \mathbf{V} \right). \tag{21}$$

Specifically, by substituting (20) and (21) into problem (18), we obtain the problem

$$\underset{\mathbf{V} \succeq \mathbf{0}}{\text{minimize}} \quad \mathrm{tr}(\mathbf{M}\mathbf{V}) \tag{22a}$$

$$\text{s.t.} \quad \mathbf{V}(m,m) = 1, \; m \in \{1, \ldots, \bar{n}_I + 1\}, \tag{22b}$$

$$\mathrm{rank}(\mathbf{V}) \leq 1, \tag{22c}$$

with the matrix $\mathbf{M}$ defined as

$$\mathbf{M} = \sum_{k \in \mathcal{N}_W} \sigma_{\theta,k}^2 \begin{bmatrix} \mathbf{a}_k \mathbf{a}_k^H & b_k \mathbf{a}_k \\ b_k^* \mathbf{a}_k^H & 0 \end{bmatrix} \tag{23}$$

$$+ \sum_{i \in \mathcal{N}_A, k \in \mathcal{N}_W} \begin{bmatrix} \mathbf{c}_{i,k} \mathbf{c}_{i,k}^H & d_{i,k} \mathbf{c}_{i,k} \\ d_{i,k}^* \mathbf{c}_{i,k}^H & 0 \end{bmatrix}.$$

To address the non-convexity of constraint (22c), we note that (22c) is equivalent to the constraint [6]

$$\mathrm{tr}(\mathbf{V}) - \sigma_1(\mathbf{V}) = 0, \tag{24}$$

where $\sigma_1(\cdot)$ denotes the largest singular value of the input matrix. Function $\sigma_1(\mathbf{V})$ is convex in $\mathbf{V}$ [18]. Furthermore, for $\mathbf{V} \succeq \mathbf{0}$, the left-hand side (LHS) of (24) is 0 when $\mathrm{rank}(\mathbf{V}) \leq 1$ and it becomes larger than 0 otherwise.

Based on this observation, as in [6], we tackle the problem

$$\underset{\mathbf{V} \succeq \mathbf{0}}{\text{minimize}} \quad \mathrm{tr}(\mathbf{M}\mathbf{V}) + \gamma (\mathrm{tr}(\mathbf{V}) - \sigma_1(\mathbf{V})) \tag{25a}$$

$$\text{s.t.} \quad \mathbf{V}(m,m) = 1, \; m \in \{1, \ldots, \bar{n}_I + 1\}, \tag{25b}$$

with a fixed weight $\gamma \geq 0$. In problem (25), we have removed the rank constraint (22c) and instead added a penalty term $\gamma(\mathrm{tr}(\mathbf{V}) - \sigma_1(\mathbf{V}))$ to the cost function that increases if (22c) is not satisfied.

The problem (25) is a difference-of-convex (DC) problem whose locally optimal solution can be efficiently found via the concave convex procedure (CCP) approach [19]. CCP solves a sequence of convex problems obtained by linearizing the terms that induce non-convexity. In the DC problem (25), the only term that induces non-convexity is $-\gamma \cdot \sigma_1(\mathbf{V})$ in the penalty term. Linearizing $-\gamma \cdot \sigma_1(\mathbf{V})$ at a reference point $\mathbf{V} = \mathbf{V}'$ yields the upper bound [6]

$$-\gamma \cdot \sigma_1(\mathbf{V}) \leq -\gamma \cdot \mathrm{tr}\left(\mathbf{V} \mathbf{u}_1(\mathbf{V}') \mathbf{u}_1(\mathbf{V}')^H \right), \tag{26}$$

Algorithm 1 CCP based algorithm for optimizing $\mathbf{V}$

1. Initialize $\mathbf{V}^{(1)}$ as (19) with arbitrary $\bar{\mathbf{v}}$ that satisfies (18b), and set $t \leftarrow 1$

2. Update $\mathbf{V}^{(t+1)}$ as a solution of the convex problem:

$$\underset{\mathbf{V} \succeq \mathbf{0}}{\text{minimize}} \quad \mathrm{tr}(\mathbf{M}\mathbf{V}) + \gamma \left(\mathrm{tr}(\mathbf{V}) - \mathrm{tr}\left(\mathbf{V}\, \mathbf{u}_1(\mathbf{V}^{(t)}) \mathbf{u}_1(\mathbf{V}^{(t)})^H \right) \right)$$

$$\text{s.t.} \quad \mathbf{V}(m,m) = 1, \; m \in \{1, \ldots, \bar{n}_I + 1\},$$

3. Stop if $||\mathbf{V}^{(t+1)} - \mathbf{V}^{(t)}||_F^2 \leq \delta$ is satisfied. Otherwise, go back to Step 2 with $t \leftarrow t + 1$.

Algorithm 2 Proposed algorithm alternately optimizing $\mathbf{v}$ and $\mathbf{f}$

1. Initialize $\mathbf{v}^{(1)}$ as arbitrary vectors that satisfy (16b).

2. Update $\mathbf{f}^{(1)}$ according to (17) with $\mathbf{v} = \mathbf{v}^{(1)}$, and set $t \leftarrow 1$.

3. For $j \in \mathcal{N}_I$, update $\mathbf{v}_j^{(t+1)}$ as the elements from $(j-1)n_I + 1$ to jn_I of the last column of the matrix $\mathbf{V}$ obtained by Algorithm 1 with $\mathbf{f} = \mathbf{f}^{(t)}$.

4. For $j \in \mathcal{N}_I$ and $m \in \{1, \ldots, n_I\}$, update $\mathbf{v}_j^{(t+1)}(m) \leftarrow \mathbf{v}_j^{(t+1)}(m)/|\mathbf{v}_j^{(t+1)}(m)|$.

5. Update $\mathbf{f}^{(t+1)}$ according to (17) with $\mathbf{v} = \mathbf{v}^{(t+1)}$.

6. Stop if $\sum_{j \in \mathcal{N}_I} ||\mathbf{v}_j^{(t+1)} - \mathbf{v}_j^{(t)}||^2 + ||\mathbf{f}^{(t+1)} - \mathbf{f}^{(t)}||^2 \leq \delta$. Otherwise, go back to Step 3 with $t \leftarrow t + 1$.

---

where $\mathbf{u}_1(\cdot)$ returns the eigenvector of the input matrix corresponding to the largest eigenvalue. The condition (26) is satisfied with equality when $\mathbf{V} = \mathbf{V}'$. The CCP based algorithm for optimizing $\mathbf{V}$ is summarized in Algorithm 1.

Overall, the proposed algorithm that alternately optimizes the IRSs' phases $\mathbf{v}$ and the linear detector $\mathbf{f}$ is detailed in Algorithm 2. In the algorithm, we initialize $\mathbf{v}$ and $\mathbf{f}$ in Steps 1-2, and update $\mathbf{v}$ for fixed $\mathbf{f}$ in Steps 3-4. In Step 4, $\mathbf{v}$ is modified only when it does not satisfy the modulus constraints (18b). In Step 5, $\mathbf{f}$ is updated for fixed $\mathbf{v}$, and we check the convergence in Step 6.

V. NUMERICAL RESULTS

In simulation, we assume that the positions of N_W workers, N_A APs and N_I IRSs are uniformly distributed in a circular area of radius 100 m. We set the variance of local parameters to $\sigma_{\theta,k}^2 = 1$ for $k \in \mathcal{N}_W$ and assume $c_0 = 20$ dB, $\eta = 3$ in the path-loss models and $\gamma = 1$ for the penalty coefficient in (25a). For all links, we consider independent Rayleigh fading channels that are distributed as $h_{d,i,k} \sim \mathcal{CN}(0,1)$, $\mathbf{g}_{i,j} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mathbf{h}_{r,j,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. We compare the performance of the proposed optimized scheme with two baseline schemes, one without IRSs and one with IRSs whose reflecting phases are randomly chosen. In all figures, we plot the normalized MSE, which is defined as the MSE $\mathrm{e}(\mathbf{v}, \mathbf{f})$ normalized by $\mathbb{E}[|\bar{\theta}|^2]$ so that it lies in the range $[0, 1]$.

In Fig. 2, we plot the average normalized MSE versus the fronthaul capacity C for an IRS-aided C-RAN system with $N_W = 10$, $N_A = 5$, $N_I = 2$, $n_I = 10$ and $P/\sigma_z^2 \in \{5, 20\}$
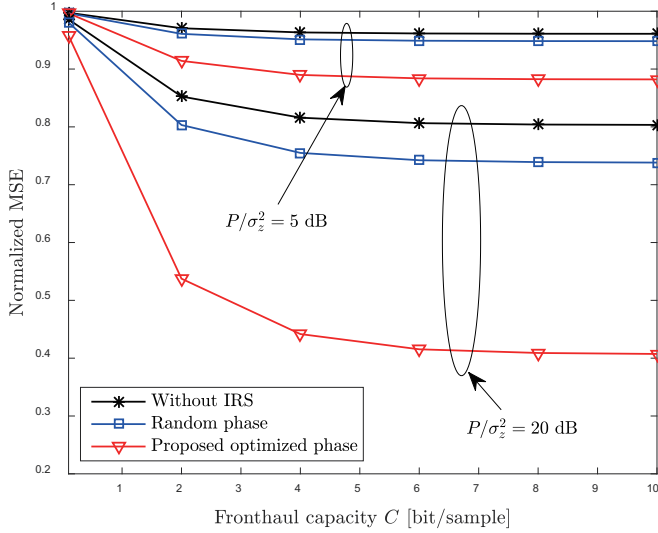
Fig. 2. Average normalized MSE versus the fronthaul capacity C for an IRS-aided C-RAN with $N_W = 10$, $N_A = 5$, $N_I = 2$, $n_I = 10$ and $P/\sigma_z^2 \in \{5, 20\}$ dB.
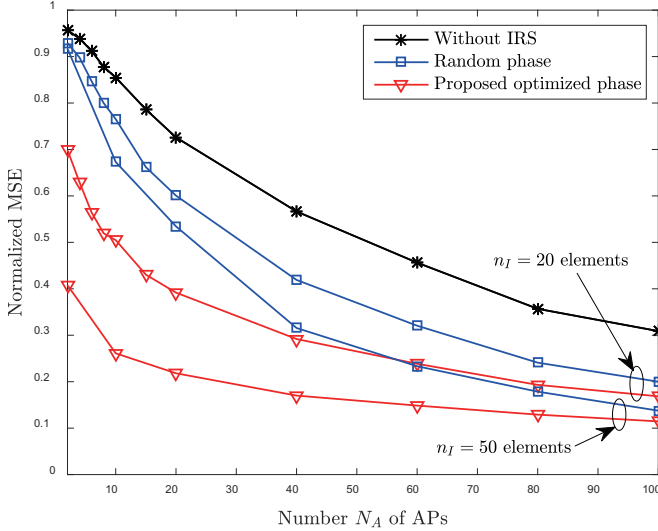

Fig. 3. Average normalized MSE versus the number N_A of APs for an IRS-aided C-RAN system with $N_W = 5$, $N_I = 2$, $n_I \in \{20, 50\}$, $C = 5$ and $P/\sigma_z^2 = 10$ dB.

dB. The figure shows that the proposed optimized scheme outperforms both baseline schemes without IRS and with random phases, and that the gain increases with the fronthaul capacity C. This is because, when C is small, the impact of carefully designing the IRSs' phases becomes minor due to the impact of the quantization noise signals $\{q_i\}_{i \in \mathcal{N}_A}$. Also, the gain increases with the signal-to-noise ratio (SNR) P/σ_z^2 of the uplink channel, and this trend coincides with the observation reported in [10, Sec. IV].

Fig. 3 plots the average normalized MSE versus the number N_A of APs for an IRS-aided C-RAN system with $N_W = 5$, $N_I = 2$, $n_I \in \{20, 50\}$, $C = 5$ and $P/\sigma_z^2 = 10$ dB. When there are only a few APs, deploying IRSs provides

relevant gains only when the reflecting phases are optimized according to Algorithm 2. However, the impact of optimizing the reflecting phases becomes minor for sufficiently large N_A.

VI. Concluding Remarks

We have studied the impacts of deploying IRSs on AirComp in a C-RAN system. To this end, we have tackled the joint optimization of the IRSs' reflecting phases and the linear detector at the PS with the goal of minimizing the MSE of the parameter estimated at the PS. Numerical results were provided that investigate the effects of various parameters on the performance gain of the proposed optimization scheme compared to baseline schemes. Among open problems, we mention the design of channel estimation process, the investigation of the effect of imperfect CSI, and the design of AirComp jointly with information transfer.

References

[1] K. Bonawitz and *et al*, "Towards federated learning at scale: System design," arXiv:1902.01046, Feb. 2019.

[2] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[3] M. M. Amiri and D. Gunduz, "Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air," in *Proc. IEEE ISIT 2019*, Paris, France, Jul. 2019, pp. 1–5.

[4] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, to appear, 2020.

[5] T. Sery and K. Cohen, "On analog gradient descent learning over multiple access fading channels," arXiv:1908.07463, Aug. 2019.

[6] T. Jiang and Y. Shi, "Over-the-air computation via intelligent reflecting surfaces," in *Proc. IEEE Globecom 2019*, Waikoloa, USA, Dec. 2019.

[7] E. Basar, M. D. Renzo, J. D. Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[8] E. Bjornson, O. Ozdogan, and E. G. Larsson, "Intelligent reflecting surface vs. decode-and-forward: How large surfaces are needed to beat relaying?," arXiv:1906.03949, Jun. 2019.

[9] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[10] C. Pan, H. Ren, K. Wang, W. Xu, A. Nallanathan M. Elkashlan, and L. Hanzos, "Multicell MIMO communications relying on intelligent reflecting surface," arXiv:1907.10864, Jul. 2019.

[11] M. Di Renzo and *et al*, "Smart radio environments empowered by reconfigurable ai meta-surfaces: an idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, pp. 1–20, May 2019.

[12] K. Ntontin and *et al*, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," arXiv:1908.08747, Aug. 2019.

[13] R. Karasik, O. Simeone, M. Di Renzo, and S. Shamai, "Beyond max-SNR: Joint encoding for reconfigurable intelligent surfaces," arXiv:1911.09443, Nov. 2019.

[14] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 69–79, 2014.

[15] X. Cao, G. Zhu, J. Xu, and K. Huang, "Optimal power control for over-the-air computation in fading channels," arXiv:1906.06858, Jun. 2019.

[16] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul. 1996.

[17] A. E. Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge Univ. Press, 2011.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, 2004.

[19] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud RAN," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, 2016.