Distributed Joint Receiver Design for Cell-Free Massive MIMO with Fast Convergence

Bikshapathi Gouda and Antti Tölli

Centre for Wireless Communications, University of Oulu, Finland Emails: {bikshapathi.gouda, antti.tolli}@oulu.fi

Abstract—Cell-Free massive MIMO system performance can be improved with cooperative beamforming strategies. Typically, local beamforming strategies are assumed at base stations (BSs) to avoid extensive channel state information exchange via backhaul links. A fully distributed framework relying on a novel overthe-air (OTA) signaling mechanism has been recently proposed to design cooperative beamformers for cell-free massive MIMO systems. However, the existing distributed exact solution has a suboptimal global convergence behaviour because of outdated information used at each BS. This paper proposes a Newton method with an adaptive regularization to design uplink receivers for the fully distributed framework. Numerical results show a significantly faster convergence of the proposed method compared to the distributed exact solution and gradient methods.

I. INTRODUCTION

Cell-free massive multiple-input multiple-output (MIMO) is one of the promising physical layer technology to improve the spectral efficiency. Cell-free massive MIMO suitably combines elements from massive MIMO, small cells, and user-centric coordinated multi-point (CoMP) with joint beamforming, to enhance overall coverage and to suppress the interference [1], [2]. Cell-free massive MIMO networks have been shown to outperform their traditional cellular massive MIMO and smallcell counterparts in several scenarios of practical interest [1], [3]. To avoid cumbersome and costly channel state information (CSI) exchange among the base stations (BSs) via backhaul signaling and to reduce the overall computational complexity, most works on cell-free massive MIMO assume simple uncoordinated beamforming strategies at the BSs, such as maximum ratio transmission/combining (MRT/MRC), local zero-forcing (ZF), and local minimum mean squared error (MMSE), which can be implemented based on local CSI. Nevertheless, the performance of cell-free massive MIMO systems can be significantly boosted by increasing the level of coordination among the BSs [3]. However, the cooperative beamforming requires enormous amount of CSI exchange via backhaul signaling, either between the BSs and a central processing unit (CPU) or among the BSs [4].

Recently, a fully distributed framework was proposed to design the cooperative beamforming strategies at each BS for cell-free massive MIMO systems [5], [6]. The performance of the fully distributed precoder design in downlink is close to the CPU based centralized design [5]. However, the fully distributed receiver design in uplink scenarios [6] has a somewhat higher performance loss with respect to the centralized design. The exact minimization used for the receiver design

in [6] turns out to be equivalent to the distributed Newton method introduced in this paper. Due to delayed over-the-air (OTA) feedback, the distributed Newton method has outdated global second order information available at each BS, which leads to overly aggressive optimization towards incorrect global direction resulting in suboptimal performance. As an alternative solution we have analyzed the distributed gradient (GD) method, which indeed provides the same performance as the centralized design, but with much slower convergence as it uses only first order information for beamformer updates.

In this paper, we propose a distributed Newton method with additional adaptive regularization term for the considered sum MSE minimization problem. In the proposed method, the outdated second order information is regulated with a variable regularization term for a faster global convergence. Numerical results show a faster convergence of the proposed method compared to the distributed GD and the distributed Newton methods, and the final converged value is very close to the centralized design.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a cell-free massive MIMO system where a set of *M*-antenna BSs $\mathcal{B} \triangleq \{1, \ldots, B\}$ serves a set of *N*antenna user equipments (UEs) $\mathcal{K} \triangleq \{1, \ldots, K\}$ in the uplink. Assuming a time division duplex (TDD) setting and a single data stream transmitted by each UE, let $\mathbf{H}_{b,k} \in \mathbb{C}^{M \times N}$ denote the uplink channel matrix between UE $k \in \mathcal{K}$ and BS $b \in \mathcal{B}$, with $\mathbf{H}_k \triangleq [\mathbf{H}_{1,k}^T, \ldots, \mathbf{H}_{B,k}^T]^T \in \mathbb{C}^{BM \times N}$ being the aggregated uplink channel matrix seen by UE k. Moreover, let $\mathbf{w}_k \in \mathbb{C}^{N \times 1}$ denote the precoding vector used by UE k. The receive signal at BS b is given by

$$\mathbf{y}_{b} \triangleq \sum_{k \in \mathcal{K}} \mathbf{H}_{b,k} \mathbf{w}_{k} d_{k} + \mathbf{z}_{b} \in \mathbb{C}^{M \times 1}$$
(1)

where $d_k \sim \mathcal{CN}(0, 1)$ is the transmit data symbol of UE k and $\mathbf{z}_b \in \mathbb{C}^{M \times 1}$ is the additive white Gaussian noise (AWGN) term at BS b, with elements distributed as $\mathcal{CN}(0, \sigma_{\text{BS}}^2)$. Likewise, the aggregated receive signal at all the BSs is given by

$$\mathbf{y} \triangleq \sum_{k \in \mathcal{K}} \mathbf{H}_k \mathbf{w}_k d_k + \mathbf{z} \in \mathbb{C}^{BM \times 1}$$
(2)

with $\mathbf{z} \triangleq [\mathbf{z}_1^{\mathrm{T}}, \dots, \mathbf{z}_B^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{BM \times 1}$. Let $\mathbf{v}_{b,k} \in \mathbb{C}^{M \times 1}$ denote the BS-specific combining vector used by BS *b* for UE *k*, with $\mathbf{v}_k \triangleq [\mathbf{v}_{1,k}^{\mathrm{T}}, \dots, \mathbf{v}_{B,k}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{BM \times 1}$ being the aggregated combining vector used for UE *k* (accordingly, we have $\sum_{b \in \mathcal{B}} \mathbf{v}_{b,k}^{\mathrm{H}} \mathbf{H}_{b,\bar{k}} = \mathbf{v}_k^{\mathrm{H}} \mathbf{H}_{\bar{k}}$). Each BS *b* estimates d_k by combining \mathbf{y}_b with $\mathbf{v}_{b,k}$ and the estimates from all the BSs

are collected by the CPU. The resulting signal-to-interferenceplus-noise ratio (SINR) for UE k reads as

$$\operatorname{SINR}_{k} \triangleq \frac{|\sum_{b \in \mathcal{B}} \mathbf{v}_{b,k}^{\mathrm{H}} \mathbf{H}_{b,k} \mathbf{w}_{k}|^{2}}{\sum_{\bar{k} \in \mathcal{K} \setminus \{k\}} |\sum_{b \in \mathcal{B}} \mathbf{v}_{b,k}^{\mathrm{H}} \mathbf{H}_{b,\bar{k}} \mathbf{w}_{\bar{k}}|^{2} + \sigma_{\operatorname{BS}}^{2} \sum_{b \in \mathcal{B}} \|\mathbf{v}_{b,k}\|^{2}}.$$
 (3)

We use the sum rate $R \triangleq \sum_{k \in \mathcal{K}} \log_2(1 + \text{SINR}_k)$ (measured in bps/Hz) as performance metric of the overall system.

In this paper, we target the sum MSE minimization problem to optimize the precoding vectors $\{\mathbf{w}_k\}_{k\in\mathcal{K}}$ and the combining vectors $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$. This can be used as a surrogate of the more involved weighted sum rate maximization problem (or, equivalently, of the iterative weighted sum MSE minimization problem [7]). In fact, since the total number of BS antennas in the network *BM* is much larger than the number of UEs *K*, the weighted sum MSE minimization yields only a minor penalty in terms of sum-rate performance as compared with the weighted sum rate maximization, while being much easier to handle and providing an inherent fairness across the UEs.

From (2), we define the MSE at UE k as

$$MSE_{k} \triangleq \mathbb{E}\left[\left|\sum_{b \in \mathcal{B}} \mathbf{v}_{b,k}^{\mathrm{H}} \mathbf{y}_{b} - d_{k}\right|^{2}\right]$$
$$= \sum_{\bar{k} \in \mathcal{K}} \left|\sum_{b \in \mathcal{B}} \mathbf{v}_{b,k}^{\mathrm{H}} \mathbf{H}_{b,\bar{k}} \mathbf{w}_{\bar{k}}\right|^{2} \qquad (4)$$
$$- 2\operatorname{Re}\left[\sum_{b \in \mathcal{B}} \mathbf{v}_{b,k}^{\mathrm{H}} \mathbf{H}_{b,k} \mathbf{w}_{k}\right] + \sigma_{\mathrm{BS}}^{2} \sum_{b \in \mathcal{B}} \|\mathbf{v}_{b,k}\|^{2} + 1$$

which is convex with respect to either the transmit or the receive strategies (but not jointly convex with respect to both). Therefore, we use alternate optimization, i.e., we optimize the combining vectors $(\mathbf{v}_{b,k})$ for fixed precoding vectors (\mathbf{w}_k) and vice versa in an iterative best-response fashion, with guaranteed local convergence [7]. Next, we describe realistic pilot-aided CSI acquisition at both the BSs and the UEs, which will be needed in Sections III and IV.

A. Uplink Pilot-Aided Channel Estimation

Let $\mathbf{h}_{b,k} \triangleq \mathbf{H}_{b,k} \mathbf{w}_k \in \mathbb{C}^{M \times 1}$ denote the effective uplink channel vector between UE k and BS b, and let $\mathbf{p}_k \in \mathbb{C}^{\tau \times 1}$ denote the pilot assigned to UE k, with $\|\mathbf{p}_k\|^2 = \tau$. In the uplink pilot-aided channel estimation phase, each UE k synchronously transmits its pilot \mathbf{p}_k precoded with \mathbf{w}_k , i.e.,

$$\mathbf{X}_{k}^{\text{UL-1}} \triangleq \mathbf{w}_{k} \mathbf{p}_{k}^{\text{H}} \in \mathbb{C}^{N \times \tau}.$$
 (5)

Then, the receive signal at BS b is given by

$$\mathbf{Y}_{b}^{\text{UL-1}} \triangleq \sum_{k \in \mathcal{K}} \mathbf{H}_{b,k} \mathbf{X}_{k}^{\text{UL-1}} + \mathbf{Z}_{b}^{\text{UL-1}}$$
(6)

$$= \sum_{k \in \mathcal{K}} \mathbf{h}_{b,k} \mathbf{p}_{k}^{\mathrm{H}} + \mathbf{Z}_{b}^{\mathrm{UL-1}} \in \mathbb{C}^{M \times \tau}$$
(7)

where $\mathbf{Z}_{b}^{\text{UL-1}}$ is the AWGN term at BS *b* with elements distributed as $\mathcal{CN}(0, \sigma_{\text{BS}}^2)$, and the least-squares (LS) estimate of $\mathbf{h}_{b,k}$ is obtained as

$$\hat{\mathbf{h}}_{b,k} \triangleq \frac{1}{\tau} \mathbf{Y}_{b}^{\text{UL-1}} \mathbf{p}_{k} \tag{8}$$

$$= \mathbf{h}_{b,k} + \frac{1}{\tau} \sum_{\bar{k} \in \mathcal{K} \setminus \{k\}} \mathbf{h}_{b,\bar{k}} \mathbf{p}_{\bar{k}}^{\mathrm{H}} \mathbf{p}_{k} + \frac{1}{\tau} \mathbf{Z}_{b}^{\mathrm{UL-1}} \mathbf{p}_{k}.$$
(9)

Here, perfect channel estimation is attained when:

i) The pilot contamination in the second term of (9) is eliminated using, e.g., orthogonal pilots (i.e., $\{\mathbf{p}_{\bar{k}}^{\mathrm{H}}\mathbf{p}_{k} = 0\}_{\bar{k}\in\mathcal{K}\setminus\{k\}}$) or non-orthogonal random pilots with infinite

length (i.e., $\tau \to \infty$);

ii) The channel estimation noise in the third term of (9) is eliminated using pilots with infinite length.

In the centralized joint receiver design, the estimation of the channel matrix $\mathbf{H}_{b,k}$ calls for N antenna-specific pilots for UE k. In this context, let $\mathbf{P}_k \in \mathbb{C}^{\tau \times N}$ denote the pilot matrix assigned to UE k, with $\|\mathbf{P}_k\|_{\mathrm{F}}^2 = \tau N$. As in (5), each UE k synchronously transmits its pilot matrix, i.e.,

$$\mathbf{X}_{k}^{\text{UL}} \triangleq \sqrt{\beta^{\text{UL}}} \mathbf{P}_{k}^{\text{H}} \in \mathbb{C}^{N \times \tau}$$
(10)

where the scaling factor $\beta^{\text{UL}} \triangleq \rho_{\text{UE}}/N$ ensures that each column of \mathbf{X}_k^{UL} complies with the UE transmit power constraint. Then, the receive signal at BS *b* is given by

$$\mathbf{Y}_{b}^{\text{UL}} \triangleq \sum_{k \in \mathcal{K}} \mathbf{H}_{b,k} \mathbf{X}_{k}^{\text{UL}} + \mathbf{Z}_{b}^{\text{UL}} \qquad (11)$$
$$= \sqrt{\beta_{\text{UL}}} \sum_{k \in \mathcal{K}} \mathbf{H}_{b,k} \mathbf{P}_{k}^{\text{H}} + \mathbf{Z}_{b}^{\text{UL}} \in \mathbb{C}^{M \times \tau} \qquad (12)$$

$$= \sqrt{\beta^{\text{UL}} \sum_{k \in \mathcal{K}} \mathbf{H}_{b,k} \mathbf{P}_k^{\text{H}} + \mathbf{Z}_b^{\text{UL}} \in \mathbb{C}^{M \times \tau}}$$
(12)

where $\mathbf{Z}_{b}^{\text{UL}}$ is the AWGN term at BS *b* with elements distributed as $\mathcal{CN}(0, \sigma_{\text{BS}}^2)$, and the LS estimate of $\mathbf{H}_{b,k}$ is obtained as

$$\hat{\mathbf{H}}_{b,k} \triangleq \frac{1}{\tau \sqrt{\beta^{\text{UL}}}} \mathbf{Y}_{b}^{\text{UL}} \mathbf{P}_{k}$$
(13)

$$= \frac{1}{\tau} \sum_{\bar{k} \in \mathcal{K}} \mathbf{H}_{b,\bar{k}} \mathbf{P}_{\bar{k}}^{\mathrm{H}} \mathbf{P}_{k} + \frac{1}{\tau \sqrt{\beta^{\mathrm{UL}}}} \mathbf{Z}_{b}^{\mathrm{UL-1}} \mathbf{P}_{k}.$$
 (14)

These observations of (9) also hold for (14), for perfect channel estimation.

B. Downlink Pilot-Aided Channel Estimation

Let $\mathbf{g}_k \triangleq \sum_{b \in \mathcal{B}} \mathbf{H}_{b,k}^{\mathrm{H}} \mathbf{v}_{b,k} \in \mathbb{C}^{N \times 1}$ denote the effective downlink channel vector between all the BSs and UE k. In the downlink pilot-aided channel estimation phase, each BS b synchronously transmits a superposition of the pilots $\{\mathbf{p}_k\}_{k \in \mathcal{K}}$ precoded with the corresponding combining vector $\mathbf{v}_{b,k}$, i.e.,

$$\mathbf{X}_{b}^{\text{DL}} \triangleq \sqrt{\beta^{\text{DL}}} \sum_{k \in \mathcal{K}} \mathbf{v}_{b,k} \mathbf{p}_{k}^{\text{H}} \in \mathbb{C}^{M \times \tau}$$
(15)

where the scaling factor β^{DL} (equal for all the BSs) ensures that each column of $\mathbf{X}_{b}^{\text{DL}}$ complies with the BS transmit power constraint. Then, the receive signal at UE k is given by

$$\mathbf{Y}_{k}^{\text{DL}} \triangleq \sum_{b \in \mathcal{B}} \mathbf{H}_{b,k}^{\text{H}} \mathbf{X}_{b}^{\text{DL}} + \mathbf{Z}_{k}^{\text{DL}}$$
(16)
$$= \sqrt{\beta^{\text{DL}}} \sum_{b \in \mathcal{B}} \sum_{\bar{k} \in \mathcal{K}} \mathbf{H}_{b,k}^{\text{H}} \mathbf{v}_{b,\bar{k}} \mathbf{p}_{\bar{k}}^{\text{H}} + \mathbf{Z}_{k}^{\text{DL}} \in \mathbb{C}^{N \times \tau}$$
(17)

where \mathbf{Z}_k^{DL} is the AWGN term at UE k with elements distributed as $\mathcal{CN}(0, \sigma_{\text{UE}}^2)$, and the LS estimate of \mathbf{g}_k is

$$\hat{\mathbf{g}}_{k} \triangleq \frac{1}{\tau \sqrt{\beta^{\text{DL}}}} \mathbf{Y}_{k}^{\text{DL}} \mathbf{p}_{k}$$
(18)

$$= \mathbf{g}_{k} + \frac{1}{\tau} \sum_{b \in \mathcal{B}} \sum_{\bar{k} \in \mathcal{K} \setminus \{k\}} \mathbf{H}_{b,k}^{\mathrm{H}} \mathbf{v}_{b,\bar{k}} \mathbf{p}_{\bar{k}}^{\mathrm{H}} \mathbf{p}_{k} + \frac{1}{\tau \sqrt{\beta^{\mathrm{DL}}}} \mathbf{Z}_{k}^{\mathrm{DL}} \mathbf{p}_{k}.$$
(19)

These observations of (9) also hold for (19), for perfect channel estimation.

III. CENTRALIZED DESIGN

In the centralized design, channel $\{\mathbf{H}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$ is estimated using the antenna specific pilots as in (13) at each BS. All the BSs feed back the estimated CSI to the CPU via backhaul links. At the CPU the estimate of the sum MSE can be written as in (20) at the top of next page. The CPU performs the alternate optimization to obtain the optimized precoder $\{\mathbf{w}_k\}_{k\in\mathcal{K}}$ and combiner $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$.

$$\sum_{k \in \mathcal{K}} \text{MSE}_{k} \simeq \sum_{k \in \mathcal{K}} \sum_{\bar{k} \in \mathcal{K}} \left| \sum_{b \in \mathcal{B}} \mathbf{v}_{b,k}^{\text{H}} \hat{\mathbf{H}}_{b,\bar{k}} \mathbf{w}_{\bar{k}} \right|^{2} - 2 \sum_{k \in \mathcal{K}} \text{Re} \left[\sum_{b \in \mathcal{B}} \mathbf{v}_{b,k}^{\text{H}} \hat{\mathbf{H}}_{b,k} \mathbf{w}_{k} \right] + \sigma_{\text{BS}}^{2} \sum_{k \in \mathcal{K}} \sum_{b \in \mathcal{B}} \|\mathbf{v}_{b,k}\|^{2} + K.$$
(20)

A. Optimization of the precoding vectors

For a fixed set of combining vectors $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$, the precoding vectors $\{\mathbf{w}_k\}_{k\in\mathcal{K}}$ are computed by minimizing the sum MSE given in (20). The precoding vector \mathbf{w}_k that minimizes (20) is the MMSE precoder

$$\mathbf{w}_{k} = \left(\sum_{\bar{k}\in\mathcal{K}} \left(\sum_{b\in\mathcal{B}} \hat{\mathbf{H}}_{b,k}^{\mathrm{H}} \mathbf{v}_{b,\bar{k}}\right) \left(\sum_{b\in\mathcal{B}} (\mathbf{v}_{b,\bar{k}}^{\mathrm{H}}) \hat{\mathbf{H}}_{b,k}\right) + \lambda_{k} \mathbf{I}_{N}\right)^{-1} \times \sum_{b\in\mathcal{B}} \hat{\mathbf{H}}_{b,k}^{\mathrm{H}} \mathbf{v}_{b,k}$$
(21)

where λ_k is the dual variable associated with the UE k power constraint. For ideal channel estimates, $\hat{\mathbf{H}}_{b,k}$ in (21) is replaced with $\mathbf{H}_{b,k}$.

B. Optimization of the combining vectors

For a fixed set of precoding vectors $\{\mathbf{w}_k\}_{k\in\mathcal{K}}$, the CPU computes the BS-specific combining vectors $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$ as a solution of sum MSE given in (20). Next, we discuss the exact minimization of (20) to obtain the BS-specific combining vectors $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$.

1) Exact Minimization

The combining vector $\{\mathbf{v}_k\}_{k\in\mathcal{K}}$ that minimizes (20) is the MMSE receiver

$$\mathbf{v}_{k} = \left(\sum_{\bar{k}\in\mathcal{K}}\hat{\mathbf{H}}_{\bar{k}}\mathbf{w}_{\bar{k}}\mathbf{w}_{\bar{k}}^{\mathrm{H}}\hat{\mathbf{H}}_{\bar{k}}^{\mathrm{H}} + \sigma_{\mathrm{Bs}}^{2}\mathbf{I}_{BM}\right)^{-1}\hat{\mathbf{H}}_{k}\mathbf{w}_{k}.$$
 (22)

For the subsequent use, we introduce the following preliminary definitions: $\mathbf{h}_k \triangleq [\mathbf{h}_{1,k}^{\mathrm{T}}, \dots, \mathbf{h}_{B,k}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{BM \times 1}$, $\hat{\mathbf{h}}_k \triangleq [\hat{\mathbf{h}}_{1,k}^{\mathrm{T}}, \dots, \hat{\mathbf{h}}_{B,k}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{BM \times 1}$, $\boldsymbol{\Phi} \triangleq \sum_{k \in \mathcal{K}} \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}} \in \mathbb{C}^{BM \times BM}$, where the latter may be rewritten as

$$\boldsymbol{\Phi} \triangleq \begin{bmatrix} \boldsymbol{\Phi}_{11} & \dots & \boldsymbol{\Phi}_{1B} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{1B}^{\mathrm{H}} & \dots & \boldsymbol{\Phi}_{BB} \end{bmatrix}$$
(23)

with $\mathbf{\Phi}_{b\bar{b}} \triangleq \sum_{k \in \mathcal{K}} \mathbf{h}_{b,k} \mathbf{h}_{\bar{b},k}^{\mathrm{H}} \in \mathbb{C}^{M \times M}$. Similarly, $\hat{\mathbf{\Phi}} \triangleq \sum_{k \in \mathcal{K}} \hat{\mathbf{h}}_{k} \hat{\mathbf{h}}_{k}^{\mathrm{H}} \in \mathbb{C}^{BM \times BM}$ and $\hat{\mathbf{\Phi}}_{b\bar{b}} \triangleq \sum_{k \in \mathcal{K}} \hat{\mathbf{h}}_{b,k} \hat{\mathbf{h}}_{\bar{b},k}^{\mathrm{H}} \in \mathbb{C}^{M \times M}$.

With the above definitions the receiver combining vector $\{\mathbf{v}_k\}_{k\in\mathcal{K}}$ can be rewritten as

$$\mathbf{v}_k = (\hat{\mathbf{\Phi}} + \sigma_{\text{BS}}^2 \mathbf{I}_{BM})^{-1} \hat{\mathbf{h}}_k.$$
 (24)

For ideal channel estimates, the combining vector \mathbf{v}_k in (24) can be obtained by replacing $\hat{\mathbf{h}}_k$ with \mathbf{h}_k and, $\hat{\boldsymbol{\Phi}}$ with $\boldsymbol{\Phi}$.

As an alternative to the exact solution, and due to practical reasons that appear useful in Section IV, an iterative solution for the receiver can be attained using either gradient or Newton approaches which are described in the following.

2) Gradient Descent Method

In GD method, the receiver combing vector $\{\mathbf{v}_k\}_{k\in\mathcal{K}}$ will be updated in the negative direction of the sum MSE gradient, evaluated at the previous iteration [8]. The gradient of the sum MSE (20) is

$$\nabla_{\mathbf{v}_{k}}\left(\sum_{k\in\mathcal{K}}\mathrm{MSE}_{k}\right) = -2(\hat{\mathbf{h}}_{k} - (\hat{\boldsymbol{\Phi}} + \sigma_{\mathrm{BS}}^{2}\mathbf{I}_{BM})\overline{\mathbf{v}}_{k}), \quad (25)$$

where $\overline{\mathbf{v}}_k$ is the combining vector of the previous iteration.

The GD update of the combining vector $\{\mathbf{v}_k\}_{k\in\mathcal{K}}$, with step-size α_c is given as [8]

$$\mathbf{v}_{k} = \overline{\mathbf{v}}_{k} + 2\alpha_{c}(\hat{\mathbf{h}}_{k} - (\hat{\mathbf{\Phi}} + \sigma_{\text{\tiny BS}}^{2}\mathbf{I}_{BM})\overline{\mathbf{v}}_{k}).$$
(26)

For ideal channel estimates, the receiver combining vector \mathbf{v}_k can be obtained by replacing $\hat{\mathbf{h}}_k$ with \mathbf{h}_k and $\hat{\boldsymbol{\Phi}}$ with $\boldsymbol{\Phi}$.

3) Newton Method

In general, GD method is slow in convergence due to the fact that it uses only first order information. To increase the convergence Newton method is preferred, in this method second order information is used for a faster convergence [8]. For the sum MSE minimization given in (20), the Newton direction for combining vector \mathbf{v}_k is given as

$$\nabla_{\mathbf{v}_{k}}^{2} \left(\sum_{k \in \mathcal{K}} \mathrm{MSE}_{k} \right)^{-1} \nabla_{\mathbf{v}_{k}} \left(\sum_{k \in \mathcal{K}} \mathrm{MSE}_{k} \right) \\ = -((\hat{\mathbf{\Phi}} + \sigma_{\mathrm{BS}}^{2} \mathbf{I}_{BM})^{-1} \hat{\mathbf{h}}_{k} - \overline{\mathbf{v}}_{k}).$$
(27)

Similar to (26), the Newton update for the combining vector $\{\mathbf{v}_k\}_{k\in\mathcal{K}}$ is given as [8]

$$\mathbf{v}_{k} = (1 - \beta_{c})\overline{\mathbf{v}}_{k} + \beta_{c}(\hat{\mathbf{\Phi}} + \sigma_{\text{BS}}^{2}\mathbf{I}_{BM})^{-1}\hat{\mathbf{h}}_{k}, \qquad (28)$$

where β_c is the Newton update step-size. For ideal channel estimates, the receiver combining vector \mathbf{v}_k can be obtained by replacing $\hat{\mathbf{h}}_k$ with \mathbf{h}_k and $\hat{\boldsymbol{\Phi}}$ with $\boldsymbol{\Phi}$.

Remark 1. The optimal combining vector \mathbf{v}_k obtained with the Newton method is the same as the exact minimization for $\beta_c = 1$.

Since the centralized design requires CSI exchange between the BSs and the CPU via backhaul links, a distributed design is preferred to avoid the CSI exchange via backhaul links. Also, the complexity of the distributed design is greatly reduced compared to the centralized design [5].

IV. DISTRIBUTED DESIGN

In the distributed joint receiver design, the alternate optimization of the combining vectors $(\mathbf{v}_{b,k})$ and the precoding vectors (\mathbf{w}_k) is carried out by means of iterative bi-directional signaling between the BSs and the UEs (see [4], [6], [9]).

A. Optimization of the precoding vectors

Let us define the component of sum MSE associated with the UE k precoding vector \mathbf{w}_k as $\underline{\text{MSE}}_k$. Each UE k computes the precoding vector $\{\mathbf{w}_k\}_{k\in\mathcal{K}}$ based on the downlink receive signal \mathbf{Y}_k^{DL} . From the receive \mathbf{Y}_k^{DL} signal, UE k associated $\underline{\text{MSE}}_k$ can be written as

$$\underline{\mathrm{MSE}}_{k} \simeq \frac{1}{\tau \beta^{\mathrm{DL}}} \mathbf{w}_{k}^{\mathrm{H}} \big(\mathbf{Y}_{k}^{\mathrm{DL}} (\mathbf{Y}_{k}^{\mathrm{DL}})^{\mathrm{H}} - \tau \sigma_{\mathrm{UE}}^{2} \mathbf{I}_{N} \big) \mathbf{w}_{k} - \frac{2}{\tau \sqrt{\beta^{\mathrm{DL}}}} \mathrm{Re}[\mathbf{w}_{k}^{\mathrm{H}} \mathbf{Y}_{k}^{\mathrm{DL}} \mathbf{p}_{k}].$$
(29)

The precoding vector \mathbf{w}_k can be computed locally at UE k as $\mathbf{w}_k = \sqrt{\beta^{\text{DL}}} (\mathbf{Y}_k^{\text{DL}} (\mathbf{Y}_k^{\text{DL}})^{\text{H}} + \tau (\lambda_k - \sigma_{\text{UE}}^2) \mathbf{I}_N)^{-1} \mathbf{Y}_k^{\text{DL}} \mathbf{p}_k$, (30) where λ_k is the dual variable associated with the UE k power constraint. The ideal precoder \mathbf{w}_k with perfect CSI can be obtained as $\tau \to \infty$.

B. Optimization of the combining vectors

For the computation of the combining vectors $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$, let us define $\mathbf{Y}^{\text{UL-1}} \triangleq [(\mathbf{Y}_1^{\text{UL-1}})^{\text{T}},\ldots,(\mathbf{Y}_K^{\text{UL-1}})^{\text{T}}]^{\text{T}} \in \mathbb{C}^{BM \times \tau}$, $\mathbf{V} \triangleq [\mathbf{v}_1,\ldots,\mathbf{v}_K] \in \mathbb{C}^{BM \times K}$ and $\mathbf{P} \triangleq [\mathbf{p}_1,\ldots,\mathbf{p}_K] \in \mathbb{C}^{\tau \times K}$. Building upon this, from the uplink receive signal $\mathbf{Y}^{\text{UL-1}}$ we can write the estimated sum MSE as

$$\sum_{k \in \mathcal{K}} \mathrm{MSE}_{k} \simeq \frac{1}{\tau} \mathrm{tr} \left(\mathbf{V}^{\mathrm{H}} \mathbf{Y}^{\mathrm{UL} \cdot 1} (\mathbf{Y}^{\mathrm{UL} \cdot 1})^{\mathrm{H}} \mathbf{V} \right) - \frac{2}{\tau} \mathrm{Re} \left[\mathrm{tr} (\mathbf{V}^{\mathrm{H}} \mathbf{Y}^{\mathrm{UL} \cdot 1} \mathbf{P}) \right] + K.$$
(31)

Each BS computes combining vectors $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$ as a solution of sum MSE given in (31). Next, we discuss the exact minimization of (31) to obtain the combining vectors $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$.

1) Distributed Exact Minimization

Each BS *b* computes the combining vector $\{\mathbf{v}_{b,k}\}_{k\in\mathcal{K}}$ by minimizing sum MSE (31) for a fixed set of precoding vectors $\{\mathbf{w}_k\}_{k\in\mathcal{K}}$ and combining vectors $\{\mathbf{v}_{\bar{b},k}\}_{\bar{b}\in\mathcal{B}\setminus\{b\},k\in\mathcal{K}}$. The local combining vector $\mathbf{v}_{b,k}^*$ that minimizes (31) reads as

$$\mathbf{v}_{b,k}^{*} = \left(\mathbf{Y}_{b}^{\text{UL-1}} (\mathbf{Y}_{b}^{\text{UL-1}})^{\text{H}}\right)^{-1} \mathbf{Y}_{b}^{\text{UL-1}} \times \left(\mathbf{p}_{k} - \sum_{\bar{b} \in \mathcal{B} \setminus \{b\}} (\mathbf{Y}_{\bar{b}}^{\text{UL-1}})^{\text{H}} \overline{\mathbf{v}}_{\bar{b},k}\right), \quad (32)$$

where $\overline{\mathbf{v}}_{\bar{b},k}$ is the previously obtained receiver combining vector obtained from BS \bar{b} . For ideal channel estimates, the local combining vector $\mathbf{v}_{b,k}^*$ can be obtained as $\tau \to \infty$ and, is given as

$$\mathbf{v}_{b,k}^* = (\boldsymbol{\Phi}_{bb} + \sigma_{\scriptscriptstyle \mathsf{BS}}^2 \mathbf{I}_M)^{-1} (\mathbf{h}_{b,k} - \boldsymbol{\xi}_{b,k})$$
(33)

where we have defined

$$\boldsymbol{\xi}_{b,k} \triangleq \sum_{\bar{b} \in \mathcal{B} \setminus \{b\}} \boldsymbol{\Phi}_{b\bar{b}} \overline{\boldsymbol{\nabla}}_{\bar{b},k}.$$
(34)

To compute the combining vector $\mathbf{v}_{b,k}^*$ at BS b, $\boldsymbol{\xi}_{b,k}$ information is required from all other BSs ($b \neq \bar{b}$). This can be obtained via backhaul links as in [4]. However, the amount of backhaul signaling required in cell-free massive MIMO grows large with the number of BSs and UEs makes the practical implementation challenging. Thanks to the OTA signaling framework proposed in [5], [6] for avoiding the extensive signal exchange via backhaul links to design the cooperative precoders/combiners.

With the OTA signaling framework, each BS *b* can obtain an estimate of $\boldsymbol{\xi}_{b,k}$ in (34) without any backhaul signaling for CSI exchange among the BSs [6]. This is achieved using an extra uplink signaling resource, whereby each UE *k* synchronously transmits $\mathbf{Y}_{k}^{\text{DL}}$ in (17) precoded with the rank-1 matrix $\mathbf{w}_{k}\mathbf{w}_{k}^{\text{H}}$, i.e.,

$$\mathbf{X}_{k}^{\text{UL-2}} \triangleq \sqrt{\beta^{\text{UL-2}}} \mathbf{w}_{k} \mathbf{w}_{k}^{\text{H}} \mathbf{Y}_{k}^{\text{DL}} \in \mathbb{C}^{N \times \tau},$$
(35)

where β^{UL-2} is the scaling factor. The scaling factor β^{UL-2} ensures that each column of \mathbf{X}_{k}^{UL-2} is transmitted with a large enough power while still complying with the UE transmit power constraint. More specifically, each UE k uses its precoding

vector \mathbf{w}_k to combine \mathbf{Y}_k^{DL} and then transmits it using again \mathbf{w}_k . Then, the receive signal at BS b is given by

$$\mathbf{Y}_{b}^{\text{UL-2}} \triangleq \sum_{k \in \mathcal{K}} \mathbf{H}_{b,k} \mathbf{X}_{k}^{\text{UL-2}} + \mathbf{Z}_{b}^{\text{UL-2}} \in \mathbb{C}^{M \times \tau}, \qquad (36)$$

where $\mathbf{Z}_{b}^{\text{UL2}} \in \mathbb{C}^{M \times \tau}$ is the AWGN term at BS *b* with elements distributed as $\mathcal{CN}(0, \sigma_{\text{BS}}^2)$. Each BS *b* can obtain an estimate of $\boldsymbol{\xi}_{b.k}$ as follows (for more details see [6])

$$\hat{\boldsymbol{\xi}}_{b,k} = \frac{1}{\tau} \left(\frac{1}{\sqrt{\beta^{\text{UL-2}}\beta^{\text{DL}}}} \mathbf{Y}_{b}^{\text{UL-2}} \mathbf{p}_{k} - \left(\mathbf{Y}_{b}^{\text{UL-1}} (\mathbf{Y}_{b}^{\text{UL-1}})^{\text{H}} - \tau \sigma_{\text{BS}}^{2} \mathbf{I}_{M} \right) \overline{\mathbf{v}}_{b,k} \right).$$
(37)

For ideal estimates of $\mathbf{Y}_{b}^{\text{UL-1}}$ and $\mathbf{Y}_{b}^{\text{UL-2}}$, above $\hat{\boldsymbol{\xi}}_{b,k}$ simplifies to $\boldsymbol{\xi}_{b,k}$ given in (34). Finally using $\mathbf{Y}_{b}^{\text{UL-2}}$ signal, $\mathbf{v}_{b,k}^{*}$ can be rewritten as

$$\mathbf{v}_{b,k}^{*} = \left(\mathbf{Y}_{b}^{\text{UL-1}}(\mathbf{Y}_{b}^{\text{UL-1}})^{\text{H}}\right)^{-1} \left(\mathbf{Y}_{b}^{\text{UL-1}}\left(\mathbf{p}_{k} + (\mathbf{Y}_{b}^{\text{UL-1}})^{\text{H}}\overline{\mathbf{v}}_{b,k}\right) - \tau \sigma_{\text{BS}}^{2}\overline{\mathbf{v}}_{b,k} - \frac{\mathbf{Y}_{b}^{\text{UL-2}}\mathbf{p}_{k}}{\sqrt{\beta^{\text{UL-2}}\beta^{\text{DL}}}}\right).$$
(38)

In parallel and distributed system, best response update on local computed values $\mathbf{v}_{b,k}^*$ leads to the global optimal solution with guaranteed convergence [10].

The best response update of the BS specific precoding vector $\mathbf{v}_{b,k}$ is performed as follows

$$\mathbf{v}_{b,k} = (1 - \gamma_d) \overline{\mathbf{v}}_{b,k} + \gamma_d \mathbf{v}_{b,k}^*, \tag{39}$$

where γ_d is the best response step-size.

Proposition 1. For $\{\overline{\mathbf{v}}_{\bar{b},k}\}_{\bar{b}\in\mathcal{B}\setminus\{b\},k\in\mathcal{K}} = \{\mathbf{v}_{\bar{b},k}\}_{\bar{b}\in\mathcal{B}\setminus\{b\},k\in\mathcal{K}}$ in (34), the combining vector $\mathbf{v}_{b,k}$ obtained with the distributed exact minimization (39) is the same as the one given by the Newton method in (28), when using ideal channel estimates and $\gamma_d = \beta_c$.

Proof. This can be proved using Schur complement, similarly to the proof given in Appendix-I of [5]. \Box

Remark 2. In practice each BS *b* obtains the combining vector $\{\mathbf{v}_{\bar{b},k}\}_{\bar{b}\in\mathcal{B}\setminus\{b\}}$ outdated by at-least one iteration. This makes the distributed exact minimization solution (39) to be different from the Newton method given in (28).

In the distributed exact minimization approach, each BS *b* optimizes the combining vector $\mathbf{v}_{b,k}$ for a fixed set of combining vectors $\{\overline{\mathbf{v}}_{\bar{b},k}\}_{\bar{b}\in\mathcal{B}\setminus\{b\}}$. This implies that each BS aggressively optimizes the combining vectors towards the local optimal with the outdated combining vectors from the other BSs ($b \neq \bar{b}$), which may lead to a suboptimal global convergence. Next, we introduce alternative receiver combining vector optimization approaches using distributed GD and distributed Newton methods.

2) Distributed Gradient Descent Method

As discussed in Section III, the receiver combing vector $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$ is updated in the negative direction of the sum MSE gradient, evaluated at the previous iteration [8]. The gradient of the sum MSE (31) is

$$\nabla_{\mathbf{v}_{b,k}} \left(\sum_{k \in \mathcal{K}} \mathrm{MSE}_k \right) = -\frac{2}{\tau} \mathbf{Y}_b^{\mathrm{UL-1}} \left(\mathbf{p}_k - \sum_{\bar{b} \in \mathcal{B}} (\mathbf{Y}_{\bar{b}}^{\mathrm{UL-1}})^{\mathrm{H}} \overline{\mathbf{v}}_{\bar{b},k} \right).$$
(40)

Similar to (26), the GD update for the combining vector $\mathbf{v}_{b,k}$ is given as

$$\mathbf{v}_{b,k} = \overline{\mathbf{v}}_{b,k} + \frac{2}{\tau} \alpha_d \mathbf{Y}_b^{\text{UL-1}} (\mathbf{p}_k - \sum_{\overline{b} \in \mathcal{B}} (\mathbf{Y}_{\overline{b}}^{\text{UL-1}})^{\text{H}} \overline{\mathbf{v}}_{\overline{b},k}), \quad (41)$$

where α_d is the gradient step-size. Using $\mathbf{Y}_b^{\text{UL-2}}$ the receiver combining vector update can be written as

$$\mathbf{v}_{b,k} = \overline{\mathbf{v}}_{b,k} + \frac{2}{\tau} \alpha_d \big(\mathbf{Y}_b^{\text{UL-1}} \mathbf{p}_k - \tau \sigma_{\text{BS}}^2 \overline{\mathbf{v}}_{b,k} - \frac{1}{\sqrt{\beta^{\text{UL-2}} \beta^{\text{DL}}}} \mathbf{Y}_b^{\text{UL-2}} \mathbf{p}_k \big)$$
(42)

The ideal combining vector $\mathbf{v}_{b,k}$ with perfect CSI can be obtained as $\tau \to \infty$.

Remark 3. The optimal combining vectors $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$ obtained with distributed GD method are the same as in the centralized GD method, assuming ideal channel estimates and $\alpha_d = \alpha_c$.

Even though the distributed GD method provides the same performance as the centralized GD method, the convergence of the GD method is fairly slow.

3) Distributed Newton Method

Newton method is preferred for a faster convergence as it utilizes also the second order information [8]. For the sum MSE minimization given in (31), the Newton direction for combining vector $\mathbf{v}_{b,k}$ is given as

$$\nabla_{\mathbf{v}_{b,k}}^{2} \left(\sum_{k \in \mathcal{K}} \mathrm{MSE}_{k} \right)^{-1} \nabla_{\mathbf{v}_{b,k}} \left(\sum_{k \in \mathcal{K}} \mathrm{MSE}_{k} \right)$$
$$= \left(\mathbf{Y}_{b}^{\mathrm{UL-1}} (\mathbf{Y}_{b}^{\mathrm{UL-1}})^{\mathrm{H}} \right)^{-1} \mathbf{Y}_{b}^{\mathrm{UL-1}} \left(\mathbf{p}_{k} - \sum_{\bar{b} \in \mathcal{B}} (\mathbf{Y}_{\bar{b}}^{\mathrm{UL-1}})^{\mathrm{H}} \overline{\mathbf{v}}_{\bar{b},k} \right).$$
(43)

Similar to (26), the Newton update for the combining vector $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$ is given as

$$\mathbf{v}_{b,k} = (1 - \beta_d) \overline{\mathbf{v}}_{b,k} + \beta_d \left(\mathbf{Y}_b^{\text{UL-1}} (\mathbf{Y}_b^{\text{UL-1}})^{\text{H}} \right)^{-1} \mathbf{Y}_b^{\text{UL-1}} \\ \times \left(\mathbf{p}_k - \sum_{\bar{b} \in \mathcal{B} \setminus \{b\}} (\mathbf{Y}_{\bar{b}}^{\text{UL-1}})^{\text{H}} \overline{\mathbf{v}}_{\bar{b},k} \right), \qquad (44)$$

where β_d is the Newton update step-size. Using $\mathbf{Y}_b^{\text{UL-2}}$ the above $\mathbf{v}_{b,k}$ can be rewritten as

$$\mathbf{v}_{b,k} = (1 - \beta_d) \overline{\mathbf{v}}_{b,k} + \beta_d \big(\mathbf{Y}_b^{\text{UL-1}} (\mathbf{Y}_b^{\text{UL-1}})^{\text{H}} \big)^{-1} \\ \times \Big(\mathbf{Y}_b^{\text{UL-1}} \big(\mathbf{p}_k + (\mathbf{Y}_b^{\text{UL-1}})^{\text{H}} \overline{\mathbf{v}}_{b,k} \big) - \tau \sigma_{\text{BS}}^2 \overline{\mathbf{v}}_{b,k} - \frac{\mathbf{Y}_b^{\text{UL-2}} \mathbf{p}_k}{\sqrt{\beta^{\text{UL-2}} \beta^{\text{DL}}}} \Big).$$
(45)

The ideal combining vector $\mathbf{v}_{b,k}$ with perfect CSI can be obtained as $\tau \to \infty$.

Remark 4. The optimal combining vectors $\{\mathbf{v}_{b,k}\}_{b\in\mathcal{B},k\in\mathcal{K}}$ obtained with the distributed Newton method are the same as those acquired using the distributed exact minimization with $\gamma_d = \beta_d$.

The performance of the distributed Newton method is equivalent to distributed exact minimization. As discussed in Section IV-B1, both methods converge fast but to a suboptimal solution.

4) Adaptive-Regularized Newton Method

As discussed earlier, the distributed GD method is slow in convergence. However, the optimal value of the distributed GD method is the same as the centralized methods, because at each BS *b* combining vector $\mathbf{v}_{b,k}$ is computed entirely based

on the previously obtained combining vectors $\{\overline{\mathbf{v}}_{\bar{b},k}\}_{\bar{b}\in\mathcal{B}\setminus\{b\}}$ from all other BSs via OTA signaling and previously computed combining vector $\{\overline{\mathbf{v}}_{b,k}\}$ at BS b itself. Distributed Newton method converges faster but to a suboptimal value due to the outdated curvature information obtained from the other BSs. This leads to overly aggressive local optimization with outdated information. Combining the advantages of distributed GD and distributed Newton, in the following, we reformulate the sum MSE minimization given in (31) by adding an extra regularization term [8]. The regularization term penalizes the aggressive optimization towards the local optimal value with restricted curvature information, while the bias in the solution due to the regularization term is compensated by considering the previously obtained combining vector $\overline{\mathbf{v}}_{b,k}$ in the regularization term itself. The modified MSE problem with the regularization term at iteration i can be written as

$$\sum_{k \in \mathcal{K}} \overline{\mathrm{MSE}}_k \triangleq \sum_{k \in \mathcal{K}} \mathrm{MSE}_k + \nu^{(i)} \sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}} \|\mathbf{v}_{b,k} - \overline{\mathbf{v}}_{b,k}\|_2^2,$$
(46)

where $\nu^{(i)}$ is the weight at iteration *i*. Furthermore, ν can be varied over iterations based on the curvature information required for the global convergence. The term $\|\mathbf{v}_{b,k} - \overline{\mathbf{v}}_{b,k}\|_2^2$ penalizes the combining vector $\mathbf{v}_{b,k}$ if the new local solution is far away from the previous global solution, this steer the solution towards the global optimal.

Similar to (45), the Newton update for the combining vector $\mathbf{v}_{b,k}$ of the modified MSE (46) at iteration *i* is given as

$$\mathbf{v}_{b,k} = (1 - \beta_d) \overline{\mathbf{v}}_{b,k} + \beta_d \big(\mathbf{Y}_b^{\text{UL-1}} (\mathbf{Y}_b^{\text{UL-1}})^{\text{H}} + \tau \nu^{(i)} \mathbf{I}_M \big)^{-1} \Big(\mathbf{Y}_b^{\text{UL-1}} \\ \times \mathbf{p}_k + \big(\mathbf{Y}_b^{\text{UL-1}} (\mathbf{Y}_b^{\text{UL-1}})^{\text{H}} - \tau (\sigma_{\text{BS}}^2 - \nu^{(i)}) \big) \overline{\mathbf{v}}_{b,k} - \frac{\mathbf{Y}_b^{\text{UL-2}} \mathbf{p}_k}{\sqrt{\beta^{\text{UL-2}} \beta^{\text{DL}}}} \Big)$$

$$(47)$$

The ideal combining vector $\mathbf{v}_{b,k}$ with perfect CSI can be obtained as $\tau \to \infty$.

For a large value of ν , the solution (47) can be approximated as distributed GD, and for a small value of ν the solution can be approximated as distributed Newton update. The first order derivative of (46) with respect to combining vector $\mathbf{v}_{b,k}$ is the same as (40). Hence the optimal value of the proposed method is the same as the centralized method.

V. NUMERICAL RESULTS

We consider a cell-free scenario where B = 25 BSs, each equipped with M = 2 antennas, are placed on a square grid with distance between neighboring BSs of 30 m and height of 10 m. Furthermore, K = 16 UEs, each equipped with N = 2 antennas, are randomly dropped in the same area. Rayleigh fading channel is generated as $vec(\mathbf{H}_{b,k}) \sim \mathcal{CN}(0, \delta_{b,k}\mathbf{I}_{MN})$, where $\delta_{b,k}$ [dB] $\triangleq -30.5 - 36.7 \log_{10}(d_{b,k})$ is the large-scale fading coefficient and $d_{b,k}$ is the distance between BS b and UE k. The maximum transmit power of the BSs and UEs in the pilot-aided channel estimation phase is $\rho_{\text{BS}} = \rho_{\text{UE}} = 30$ dBm. Lastly, the AWGN power at the BSs and at the UEs is fixed to $\sigma_{\text{BS}}^2 = \sigma_{\text{UE}}^2 = -95$ dBm. In the following sum MSE and sum rate performances of the proposed OTA distributed adaptive-regularized Newton (Dis-ARNW) method are compared with the centralized joint receiver design (Cen)



Figure 1: Average sum MSE vs. OTA iterations with ideal CSI.



Figure 2: Average sum rate vs. OTA iterations with ideal CSI.

described in Section III, OTA distributed joint receiver design methods of GD (Dis-GD) and Newton (Dis-NW) described in Section IV.

In Figure 1, the average sum MSE of different approaches are compared. 'Dis-GD' converges to the 'Cen' since both methods are equivalent as shown in Section IV. 'Dis-ARNW' shows slower convergence compared to 'Dis-NW' in the initial iteration, but the global convergence of 'Dis-ARNW' is much better than that of 'Dis-NW'. Moreover, 'Dis-ARNW' typically converges close to the optimal value within 10 iteration, whereas 'Dis-GD' needs more than 20 iterations. Similar behaviour is reflected in the sum-rate plot shown in Figure 2. Corresponding behaviour of sum MSE and sum rate with the estimated channels are shown in Figures 3 and 4, respectively. However, the centralized design methods are degraded more compared to the distributed design methods, due to the fact that centralized designs use single channel estimate. Whereas distributed designs uses multiple channel estimates (OTA training), and the AWGN noise differs for each channel estimate [5].

VI. CONCLUSIONS

In this paper, we have proposed an adaptive-regularized Newton method for receiver combining vector design for the distributed system with OTA framework. The proposed method shows better convergence compared to the distributed gradient and distributed Newton methods. We have also shown that, for the considered sum MSE minimization problem, centralized GD method is the same as the distributed GD method, and



Figure 3: Average sum MSE vs. OTA iterations with estimated CSI.



Figure 4: Average sum rate vs. OTA iterations with estimated CSI.

distributed exact minimization with the best-response update is equivalent to the distributed Newton update.

REFERENCES

- H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
- [2] E. Interdonato, G. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. and Networking*, vol. 2019, no. 1, pp. 197–209, Aug. 2019.
- [3] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2020.
- [4] J. Kaleva, A. Tölli, M. Juntti, R. A. Berry, and M. L. Honig, "Decentralized joint precoding with pilot-aided beamformer estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2330–2341, May 2018.
- [5] I. Atzeni, B. Gouda, and A. Tölli, "Distributed precoding design via over-the-air signaling for cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2020.
- [6] —, "Distributed joint receiver design for uplink cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, June 2020.
- [7] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sept. 2011.
- [8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [9] A. Tölli *et al.*, "Distributed coordinated transmission with forwardbackward training for 5G radio access," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 58–64, Jan. 2019.
- [10] F. Scutari, G. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multiagent systems," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 641–656, Feb. 2014.