Fast-Convergent Distributed Coordinated Precoding for TDD Multicell MIMO Systems

Rasmus Brandt[1] and Mats Bengtsson[2]
Dept. of Signal Processing, ACCESS Linnæus Centre, KTH Royal Institute of Technology, Sweden
[1]rabr5411@kth.se, [2]mats.bengtsson@ee.kth.se

***Abstract*—Several distributed coordinated precoding methods relying on over-the-air (OTA) iterations in time-division duplex (TDD) networks have recently been proposed. Each OTA iteration incurs overhead, which reduces the time available for data transmission. In this work, we therefore propose an algorithm which reaches good sum rate performance within just a few number of OTA iterations, partially due to non-overhead-incurring local iterations at the receivers. We formulate a scalarized multi-objective optimization problem where a linear combination of the weighted sum rate and the multiplexing gain is maximized. Using a well-known heuristic for smoothing the optimization problem together with a linearization step, the distributed algorithm is derived. When numerically compared to the state-of-the-art in a scenario with 1 to 3 OTA iterations allowed, the algorithm shows significant sum rate gains at high signal-to-noise ratios.**

I. Introduction

Coordinated precoding [1] is a promising technique for improving the data rates in future 5G wireless networks. Its implementation typically requires a large amount of channel state information (CSI) at the nodes of the network. Lately, several works (see [2] and references therein) have studied local CSI acquisition at the transmitters by exploiting the reciprocity of the channel when time-division duplex (TDD) is used. In this mode of operation, each node in the network can perform one optimization iteration based on its current knowledge of local effective channels [2]. This enables iterative coordinated precoding algorithms (see [3] and references therein) to be implemented in a fully distributed manner.

Depending on the deployment scenario, the number of such over-the-air (OTA) iterations may be constrained however. This may be due to short coherence times of the channel or due to the design of the uplink/downlink switching periodicity of the frame structure[1]. In this work, we therefore develop a distributed coordinated precoding algorithm which reaches high sum rates within just a couple of OTA iterations. The algorithm is developed by first formulating a novel optimization problem using notions from interference alignment (IA) [5] and insights from existing iterative coordinated precoding algorithms [3]. The resulting multiplexing gain-regularized weighted sum rate optimization problem is approximated using a well-known log-det heuristic from the sparse signal processing literature [6] in order to handle the non-smooth multiplexing gain term. A stationary point of an approximated problem is found by exploiting the relationship between the rates and the minimum mean squared errors [7] and applying block coordinate descent [8] to a linearized version of the approximated problem. The resulting algorithm has fast convergence, partially due to non-overhead-incurring local iterations at the receivers. Numerical performance evaluation shows a significant sum rate improvement of the algorithm, when compared to state-of-the-art in an over-the-air iteration constrained scenario.

Existing works on OTA constrained coordinated precoding includes [9], which however treats the leakage minimization problem in interference alignment. The multiplexing gain was previously heuristically approximated in [10], where a nuclear norm heuristic was used, and in [11] where a weighted nuclear norm was used. Contrary to this work however, the impact of the desired effective channel on the rates is not directly taken into account in the algorithms proposed in [10], [11].

[1]In TD-LTE, the minimum interval between OTA iterations is 5 ms [4], thus constraining the maximum number of OTA iterations per coherence interval.

II. Problem Formulation

The multicell MIMO network considered is modelled as an interfering broadcast channel with I base stations (BSs). BS i serves K_i mobile stations (MSs) using coordinated precoding[2]. We index the kth MS connected to the ith BS with the index pair (i,k). For brevity, we will often write this index pair as i_k. We denote the $N_{i_k} \times M_j$ MIMO channel from BS j to MS i_k as $\mathbf{H}_{i_k j}$. The transmitted signal intended for MS i_k is $\mathbf{x}_{i_k} \in \mathbb{C}^{d_{i_k}}$ where $d_{i_k} \in \mathbb{Z}^+$ is the number of data streams transmitted to that user. Specifically, we assume that a linear precoder $\mathbf{V}_{i_k} \in \mathbb{C}^{M_i \times d_{i_k}}$ is used at the transmitter. With additive white Gaussian noise $\mathbf{n}_{i_k} \sim \mathcal{CN}\left(\mathbf{0}, \sigma_{i_k}^2 \mathbf{I}\right)$, the signal model for the received signal at MS i_k is

$$\mathbf{s}_{i_k} = \mathbf{H}_{i_k i}\mathbf{V}_{i_k}\mathbf{x}_{i_k} + \sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{x}_{j_l} + \mathbf{n}_{i_k},$$

where the second term constitutes the sum of inter-cell and intra-cell interference. The performance metric is the achievable rate[3], which for MS i_k is given by $r_{i_k} \triangleq \log\det\left(\mathbf{I} + \mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\left(\mathbf{\Psi}_{i_k} + \sigma_{i_k}^2\mathbf{I}\right)^{-1}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}\right)$, where $\mathbf{\Psi}_{i_k} \triangleq \sum_{(j,l)\neq(i,k)} \mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_k j}^{\mathsf{H}}$ is the interference covariance matrix. We also define $\mathbf{\Phi}_{i_k} \triangleq \sum_{(j,l)} \mathbf{H}_{i_k j}\mathbf{V}_{j_l}\mathbf{V}_{j_l}^{\mathsf{H}}\mathbf{H}_{i_k j}^{\mathsf{H}} + \sigma_{i_k}^2\mathbf{I}$ as the total received signal covariance for MS i_k.

A. Weighted Sum Rate Maximization and OTA Overhead

Given user priorities $\alpha_{i_k} \geq 0$, and assuming per-BS sum power constraints, a weighted sum rate (WSR) maximization

[2]Coordinated precoding means that the BSs coordinate their selection of the precoders, but they do not share user data.

[3]Here we assume perfect CSI at the receivers, Gaussian signals, long codewords, and an optimal decoder which treats interference as noise.

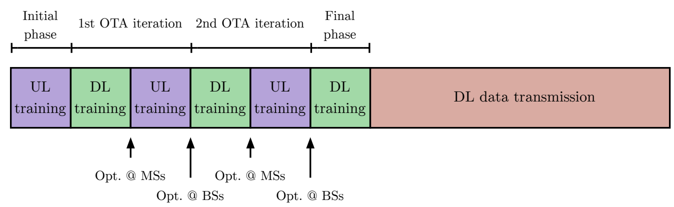


Fig. 1. Example of radio frame including $L_{\text{OTA}} = 2$ over-the-air iterations. Before the first OTA iteration, the BSs acquire local CSI from an initial uplink training phase. After the last OTA iteration, the MSs acquire local CSI for the final effective channels.

problem can be formulated:

$$\underset{\{\mathbf{V}_{i_k}\}}{\text{maximize}} \sum_{(i,k)} \alpha_{i_k} r_{i_k}, \text{ s.t.} \sum_{k=1}^{K_i} \|\mathbf{V}_{i_k}\|_{\text{F}}^2 \leq P_i, \;\; i = 1, \ldots, I. \tag{1}$$

This problem has been shown to be NP-hard, but several distributed and iterative algorithms for finding a local optimum exist, requiring only local CSI in each OTA iteration [3]. The local CSI in the downlink can be distributedly estimated using pilot-assisted channel training [2]. Under TDD mode, assuming appropriately calibrated RF chains, the corresponding local CSI at the transmitters can also be obtained distributedly by pilot-assisted channel training in the uplink [2]. Given such a training setup, the coordinated precoding optimization algorithms can distributedly perform one optimization iteration per uplink/downlink training phase; this is our definition of an OTA iteration. In Fig. 1, we give an example on how a radio frame could look like for the case of $L_{\text{OTA}} = 2$ OTA iterations before downlink data transmission.

Each OTA iteration leads to overhead since a fraction of the coherence time is used for optimization rather than data transmission. Since we focus on the optimization modelling and algorithm development, we crudely measure this overhead in terms of the number of OTA iterations used. In order to find a method that reaches high weighted sum rates in a few number of OTA iterations, our first step will be in formulating another optimization problem to be solved instead of the one in (1).

B. WSR Maximization with MG Regularization

The *multiplexing gain* (MG) of MS i_k is $\mathtt{MG}_{i_k}$ such that $r_{i_k} = \mathtt{MG}_{i_k} \log(\mathtt{SNR}_{i_k}) + o(\log(\mathtt{SNR}_{i_k}))$, where $\mathtt{SNR}_{i_k}$ is the corresponding signal-to-noise ratio (SNR). At high SNRs, it is imperative to achieve high MGs in order to achieve high rates. To explicitly describe the MG, we introduce the linear receive filter $\mathbf{U}_{i_k} \in \mathbb{C}^{N_{i_k} \times d_{i_k}}$. Inspired by [10], we then define the MG achieved under interference alignment [5] for MS i_k as

$$\mathtt{MG}_{i_k} \triangleq \text{rank}\big(\underbrace{\mathbf{U}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{U}_{i_k}}_{\triangleq \boldsymbol{\Xi}_{i_k} \in \mathbb{C}^{d_{i_k} \times d_{i_k}}}\big) - \text{rank}\left(\mathbf{U}_{i_k}^{\mathsf{H}} \boldsymbol{\Psi}_{i_k} \mathbf{U}_{i_k}\right).$$

In order to reach a large MG, the effective signal covariance matrix $\boldsymbol{\Xi}_{i_k}$ should be full rank and the interference covariance matrix $\boldsymbol{\Psi}_{i_k}$ should be rank deficient.

Instead of only optimizing the weighted sum rate as in (1), or only optimizing the sum MG as in [10], [11], we now consider a scalarized multi-objective optimization approach where we optimize a linear combination of the two. This formulation is inspired by the intuition that when iterative optimization methods are applied, the solution should be driven towards a point which is good both in terms of sum rate as well as in terms of sum MG. Given a weight $\rho \geq 0$, we thus propose the following novel optimization problem:

$$\begin{aligned} \underset{\{\mathbf{U}_{i_k}, \mathbf{V}_{i_k}\}}{\text{maximize}} \quad & \sum_{(i,k)} \alpha_{i_k} \left(r_{i_k} + \rho\, \mathtt{MG}_{i_k}\right) \\ \text{subject to} \quad & \sum_{k=1}^{K_i} \|\mathbf{V}_{i_k}\|_{\text{F}}^2 \leq P_i, \quad i = 1, \ldots, I. \end{aligned} \tag{2}$$

This optimization problem is both non-convex and non-smooth however, making it hard to solve to global optimality. We will therefore solve an approximated version of the problem.

III. The FastDCP Algorithm

We now propose a heuristic for approximating the optimization problem in (2), as well as a distributed method for finding stationary points to the approximated problem.

A. Approximated Optimization Problem

First, for tractability, we neglect the influence of the optimization variables on $\text{rank}\left(\boldsymbol{\Xi}_{i_k}\right)$, giving the rough bound $\mathtt{MG}_{i_k} \geq -\text{rank}\left(\mathbf{U}_{i_k}^{\mathsf{H}} \boldsymbol{\Psi}_{i_k} \mathbf{U}_{i_k}\right) \triangleq \overline{\mathtt{MG}}_{i_k}$. In [10], [11], $\boldsymbol{\Xi}_{i_k}$ is constrained to always be full rank. This is not the case here, but the trivial solution is avoided due to r_{i_k} being present in the objective of (2).

Next, we note that the rank of a positive semidefinite matrix $\mathbf{A}$ can be approximated as $\text{rank}\left(\mathbf{A}\right) \approx \log\det\left(\delta \mathbf{I} + \mathbf{A}\right)$ [6]. This is a smooth approximation of the discontinuous rank operator, where $\delta > 0$ determines the value of the approximation as $\mathbf{A}$ approaches the zero matrix. We thus get

$$\overline{\mathtt{MG}}_{i_k} \approx -\log\det\left(\delta \mathbf{I} + \mathbf{U}_{i_k}^{\mathsf{H}} \boldsymbol{\Psi}_{i_k} \mathbf{U}_{i_k}\right) = -\log\det\left(\mathbf{F}_{i_k}\right) = \widetilde{\mathtt{MG}}_{i_k}$$

where $\mathbf{F}_{i_k} \triangleq \delta \mathbf{I} + \mathbf{U}_{i_k}^{\mathsf{H}} \boldsymbol{\Psi}_{i_k} \mathbf{U}_{i_k}$. The resulting heuristically approximated optimization problem is then

$$\begin{aligned} \underset{\{\mathbf{U}_{i_k}, \mathbf{V}_{i_k}\}}{\text{maximize}} \quad & \sum_{(i,k)} \alpha_{i_k} \left(r_{i_k} + \rho\, \widetilde{\mathtt{MG}}_{i_k}\right) \\ \text{subject to} \quad & \sum_{k=1}^{K_i} \|\mathbf{V}_{i_k}\|_{\text{F}}^2 \leq P_i, \quad i = 1, \ldots, I. \end{aligned} \tag{3}$$

The approximated problem is smooth, but still non-convex.

It is well-known [7] that given the mean squared error (MSE) of MS i_k, $\mathbf{E}_{i_k} \triangleq \mathbf{I} - \mathbf{U}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} - \mathbf{V}_{i_k}^{\mathsf{H}} \mathbf{H}_{i_k i}^{\mathsf{H}} \mathbf{U}_{i_k} + \mathbf{U}_{i_k}^{\mathsf{H}} \boldsymbol{\Phi}_{i_k} \mathbf{U}_{i_k}$, the corresponding rate can be written as $r_{i_k} = -\inf_{\mathbf{U}_{i_k}} \log\det\left(\mathbf{E}_{i_k}\right)$ [7]. The optimization problem in (3) is therefore reformulated as

$$\begin{aligned} \underset{\{\mathbf{U}_{i_k}, \mathbf{V}_{i_k}\}}{\text{minimize}} \quad & \sum_{(i,k)} \alpha_{i_k} \left[\log\det\left(\mathbf{E}_{i_k}\right) + \rho \log\det\left(\mathbf{F}_{i_k}\right)\right] \\ \text{subject to} \quad & \sum_{k=1}^{K_i} \|\mathbf{V}_{i_k}\|_{\text{F}}^2 \leq P_i, \quad i = 1, \ldots, I. \end{aligned} \tag{4}$$

The terms in the objective of (4) are non-convex in the optimization variables. However, the first term is concave in $\mathbf{E}_{i_k}$ and the second term is concave in $\mathbf{F}_{i_k}$. As functions of these variables, the terms can thus be globally upper bounded by first-order Taylor approximations. The matrix inverses of the Taylor linearization points can then be introduced as optimization variables $\{\mathbf{Y}_{i_k}\}$ and $\{\mathbf{Z}_{i_k}\}$ (see e.g. [7] or [2]), giving a linearized and extended problem:

$$\begin{aligned}
\underset{\left\{\substack{\mathbf{U}_{i_k},\mathbf{V}_{i_k},\\ \mathbf{Y}_{i_k},\mathbf{Z}_{i_k}}\right\}}{\text{minimize}} \quad & \sum_{(i,k)} \alpha_{i_k} \big[\mathrm{Tr}\left(\mathbf{Y}_{i_k}\mathbf{E}_{i_k}\right) - \log\det\left(\mathbf{Y}_{i_k}\right) \\
& + \rho \left(\mathrm{Tr}\left(\mathbf{Z}_{i_k}\mathbf{F}_{i_k}\right) - \log\det\left(\mathbf{Z}_{i_k}\right)\right)\big] \\
\text{subject to} \quad & \sum_{k=1}^{K_i} \|\mathbf{V}_{i_k}\|_{\mathrm{F}}^2 \le P_i, \quad i = 1,\ldots,I. \qquad (5)
\end{aligned}$$

This problem is still non-convex, but the key difference to (4) is that (5) is convex in each block of variables, when the remaining three blocks are kept fixed. Due to this property, the problem lends itself to block coordinate descent [8].

For $\rho = 0$, the optimization problem in (5) is the same as the formulation in [7]. For $\rho > 0$ however, a major difference between (5) and the formulation in [7] is that (5) will allow for *local iterations* at the MSs, when it is solved through block coordinate descent. The local iterations will monotonically improve performance for the MSs, and they can be performed without requiring overhead-incurring OTA iterations. The local iterations essentially trade faster convergence for a slightly higher computational complexity, a trade which could be very favourable in OTA iteration constrained scenarios.

B. Optimality Conditions

The optimization problem in (5) has four blocks of variables: receive filters $\{\mathbf{U}_{i_k}\}$, precoders $\{\mathbf{V}_{i_k}\}$, and linearization weights $\{\mathbf{Y}_{i_k}\}$ and $\{\mathbf{Z}_{i_k}\}$. Block coordinate descent amounts to fixing all blocks of optimization variables, except one, and then optimizing with respect to that block. We will now detail the individual optimality conditions for each block.

1) Linearization Weights: For the linearization weights $\{\mathbf{Y}_{i_k}\}$ and $\{\mathbf{Z}_{i_k}\}$, it can easily be shown that the optimality conditions are $\mathbf{Y}_{i_k}^\star = \mathbf{E}_{i_k}^{-1}$ and $\mathbf{Z}_{i_k}^\star = \mathbf{F}_{i_k}^{-1}$ for all i_k.

2) Receive Filters: Fixing all blocks except the receive filters $\{\mathbf{U}_{i_k}\}$, the resulting optimization problem becomes trivially distributed over the MSs. MS i_k should therefore solve

$$\underset{\mathbf{U}_{i_k}}{\text{minimize}}\ \mathrm{Tr}\left(\mathbf{Y}_{i_k}\mathbf{E}_{i_k}\right) + \rho\, \mathrm{Tr}\left(\mathbf{Z}_{i_k}\mathbf{F}_{i_k}\right)$$

This is an unconstrained optimization problem, and the optimality condition can be shown to be

$$\rho \mathbf{\Phi}_{i_k}^{-1}\mathbf{\Psi}_{i_k}\mathbf{U}_{i_k}^\star + \mathbf{U}_{i_k}^\star\mathbf{Y}_{i_k}\mathbf{Z}_{i_k}^{-1} = \mathbf{\Phi}_{i_k}^{-1}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}\mathbf{Y}_{i_k}\mathbf{Z}_{i_k}^{-1}. \quad (6)$$

This is a Sylvester equation in $\mathbf{U}_{i_k}^\star$, for which there exist several solution algorithms; see e.g. [12] or Appendix A.

3) Precoders: Fixing all blocks except the precoders $\{\mathbf{V}_{i_k}\}$, the resulting optimization problem becomes trivially distributed over the BSs. Defining $\mathbf{\Gamma}_i \triangleq \sum_{(j,l)} \alpha_{j_l}\mathbf{H}_{j_l i}^{\mathsf{H}}\mathbf{U}_{j_l}\mathbf{Y}_{j_l}\mathbf{U}_{j_l}^{\mathsf{H}}\mathbf{H}_{j_l i}$ as the uplink total covariance, and $\mathbf{\Lambda}_{i_k} \triangleq \sum_{(j,l)\neq(i,k)} \alpha_{j_l}\mathbf{H}_{j_l i}^{\mathsf{H}}\mathbf{U}_{j_l}\mathbf{Z}_{j_l}\mathbf{U}_{j_l}^{\mathsf{H}}\mathbf{H}_{j_l i}$ as an uplink interference covariance, it can be shown that BS i should solve

$$\begin{aligned}
\underset{\{\mathbf{V}_{i_k}\}_{k=1}^{K_i}}{\text{minimize}} \quad & \sum_{k=1}^{K_i} \big[\mathrm{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}}\left(\mathbf{\Gamma}_i + \rho\,\mathbf{\Lambda}_{i_k}\right)\mathbf{V}_{i_k}\right) \\
& - \alpha_{i_k}\left(\mathrm{Tr}\left(\mathbf{U}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}\mathbf{V}_{i_k}\right) + \mathrm{Tr}\left(\mathbf{V}_{i_k}^{\mathsf{H}}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{U}_{i_k}\right)\right)\big] \\
\text{subject to} \quad & \sum_{k=1}^{K_i} \|\mathbf{V}_{i_k}\|_{\mathrm{F}}^2 \le P_i. \qquad (7)
\end{aligned}$$

For this convex optimization problem with non-empty relative interior, Slater's constraint qualification gives that strong duality holds, and we therefore solve it using the Karush-Kuhn-Tucker conditions. The stationarity condition gives that

$$\mathbf{V}_{i_k}^\star = \alpha_{i_k}\left(\mathbf{\Gamma}_i + \rho\mathbf{\Lambda}_{i_k} + \mu_i^\star\mathbf{I}\right)^{-1}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{U}_{i_k}\mathbf{Y}_{i_k},\ \forall\, i_k.$$

If $\sum_{k=1}^{K_i}\left\|\mathbf{V}_{i_k}^\star\right\|_{\mathrm{F}}^2 \le P_i$ for $\mu_i^\star = 0$, the solution has been found. Otherwise, $\mu_i^\star > 0$ is found such that $\sum_{k=1}^{K_i}\left\|\mathbf{V}_{i_k}^\star\right\|_{\mathrm{F}}^2 = P_i$. This can be done efficiently by bisection (see Appendix B).

Algorithm 1 Fast-Convergent Dist. Coord. Prec. (FastDCP)

Parameters: $\{\alpha_{i_k}\}, \rho, \delta, L_{\text{local}}, L_{\text{OTA}}$
Initialization: $\{\mathbf{V}_{i_k}\}$
repeat
  In parallel over MSs:
  repeat
    Solve (6), yielding $\mathbf{U}_{i_k}$
    Let $\mathbf{Y}_{i_k} = \mathbf{E}_{i_k}^{-1}$ and $\mathbf{Z}_{i_k} = \mathbf{F}_{i_k}^{-1}$
  until L_{local} local iterations have been performed
  In parallel over BSs:
  Find $0 \le \mu_i \le \mu_i^{\max}$ such that $\sum_{k=1}^{K_i}\|\mathbf{V}_{i_k}\|_{\mathrm{F}}^2 \le P_i$.
  $\mathbf{V}_{i_k} = \alpha_{i_k}\left(\mathbf{\Gamma}_i + \rho\mathbf{\Lambda}_{i_k} + \mu_i\,\mathbf{I}\right)^{-1}\mathbf{H}_{i_k i}^{\mathsf{H}}\mathbf{U}_{i_k}\mathbf{Y}_{i_k}$
until L_{OTA} over-the-air iterations have been performed

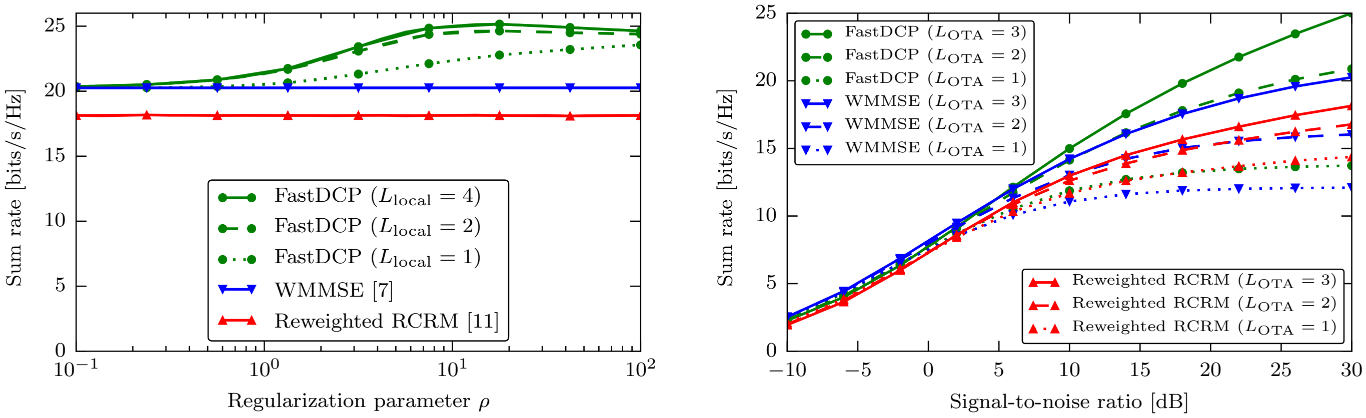
C. Iterative and Distributed Algorithm

The final *fast-convergent distributed coordinated precoding* (FastDCP) algorithm is now presented in Algorithm 1. At the MSs, given local CSI[4], L_{local} local iterations are performed per OTA iteration. Each local iteration amounts to solving (6) and updating the linearization weights. The local iterations at the MSs are possible due to the coupling between $\mathbf{U}_{i_k}$, $\mathbf{Y}_{i_k}$, and $\mathbf{Z}_{i_k}$ in the optimality conditions. This is in contrast to [7], where the receive filter does not depend on the linearization weight, leaving no opportunity for local iterations. At the BSs, given local information about channels and linearization weights[4], the optimization problem in (7) is solved. The updates are iteratively performed until L_{OTA} OTA iterations have been performed (cf. Fig. 1). We treat ρ as a fixed parameter, but it could also be adapted to the scenario circumstances.

Proposition 1. *With unbounded L_{OTA} and bounded L_{local}, any limit point $\{\mathbf{U}_{i_k}^\star, \mathbf{V}_{i_k}^\star\}$ of the FastDCP algorithm is a stationary point of the optimization problem in* (4).

Proof: The result is based on the theory in [8]. Since it can be shown that each subproblem admits a unique solution

[4]Local CSI at the receivers can be obtained through downlink channel training. Local CSI and linearization weights at the transmitters can be obtained through (reciprocal) uplink channel training and feedback [2]. These OTA iterations implicitly coordinate the precoders/receive filters in the network and since explicit information sharing is not required, a backhaul is not needed.



(a) Varying ρ and L_{local} for FastDCP. All algorithms used $L_{\text{OTA}} = 3$ and we let $\texttt{SNR} = 30\,\text{dB}$.

(b) Varying $\texttt{SNR}$ and L_{OTA} for all algorithms. FastDCP used $L_{\text{local}} = 4$ and $\rho = 10$.

Fig. 2. Performance comparison between FastDCP and the benchmarks in the $(2 \times 3, 1)^6$ interference channel.

and that the linearization between (4) and (5) satisfies the conditions of Assumption 2 in [8] (see [8, Sec. VIII-A]), the convergence to a stationary point is given by Cor. 2 of [8]. ∎

IV. Numerical Results

We study the sum rate performance of the system by means of numerical simulations. Our benchmarks are the (distributed) WMMSE algorithm from [7] and the (centralized) reweighted rank-constrained rank minimization (RCRM) heuristic from [11]. For compatibility with the reweighted RCRM, we study an interference channel where $I = 6$ BSs are serving one MS each (i.e. $K_i = 1$). The BSs have $M = 3$ antennas each and the MSs have $N = 2$ antennas each. We let $\alpha = 1$ and $d = 1$ for all MSs; thus, the scenario is not IA feasible. Note however that the reweighted RCRM in [11] (contrary to the regular RCRM in [10]) was developed to work in IA infeasible settings. The channels were i.i.d. Rayleigh fading, such that $[\mathbf{H}_{i_k j}]_{nm} \sim \mathcal{CN}(0, 1)$. We define $\texttt{SNR} = \frac{P}{\sigma^2}$, where P is the transmit power of the BSs and σ^2 is the noise power of the MSs. All algorithms were initialized with the largest right singular vector of $\mathbf{H}_{i_k i}$, and for the FastDCP algorithm, we let $\delta = 1$. The results were averaged over 200 independent Monte Carlo realizations. The source code is made available at [13].

In Fig. 2a, we show performance of the FastDCP algorithm when varying the regularization parameter ρ and the number of local iterations L_{local} while keeping $L_{\text{OTA}} = 3$ and $\texttt{SNR} = 30\,\text{dB}$ fixed. Performance is best around $\rho = 10$. As expected, the FastDCP algorithm performs similar to the WMMSE algorithm as $\rho \to 0$. The effectiveness of the local iterations is also visible. In Fig. 2b, we compare performance between the algorithms while varying the SNR and the number of OTA iterations L_{OTA} while keeping $L_{\text{local}} = 4$ and $\rho = 10$ fixed for the FastDCP algorithm. In the high-SNR regime, the reweighted RCRM algorithm is marginally the best scheme for $L_{\text{OTA}} = 1$, but for $L_{\text{OTA}} \in \{2, 3\}$, the FastDCP algorithm outperforms both benchmarks.

V. Conclusions

By jointly optimizing the sum rate and the sum multiplexing gain, a distributed and iterative algorithm can be derived which reaches high sum rate performance within just a couple of OTA iterations. This property makes the algorithm interesting for future 5G systems where a limited number of OTA iterations are part of the frame structure (e.g. like Fig. 1).

Acknowledgements

The first author would like to thank Mr. Hadi Ghauch for stimulating discussions and insightful comments on this work.

References

[1] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2-3, pp. 113–381, 2013.

[2] R. Brandt and M. Bengtsson, "Distributed CSI acquisition and coordinated precoding for TDD multicell MIMO systems," *IEEE Trans. Veh. Technol.*, 2015, accepted. http://dx.doi.org/10.1109/TVT.2015.2432051.

[3] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Comparison of distributed beamforming algorithms for MIMO interference networks," *IEEE TSP*, vol. 61, no. 13, pp. 3476–3489, 2013.

[4] E. Dahlman, S. Parkvall, and J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 2011.

[5] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, 2008.

[6] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proc. ACC'03*, vol. 3, Jun. 2003, pp. 2156–2162.

[7] Q. Shi, M. Razavivayn, Z. 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, 2011.

[8] M. Razaviyayn, M. Hong, and Z. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[9] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Distributed low-overhead schemes for multi-stream MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1737–1749, Apr. 2015.

[10] D. S. Papailiopoulos and A. G. Dimakis, "Interference alignment as a rank constrained rank minimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4278–4288, 2012.

[11] H. Du, T. Ratnarajah, M. Sellathurai, and C. B. Papadias, "Reweighted nuclear norm approach for interference alignment," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3754–3765, 2013.

[12] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation AX + XB = C," *Commun. ACM*, vol. 15, no. 9, pp. 820–826, Sep. 1972.

[13] R. Brandt, "FastDCP simulation environment," accessible at: github.com/rasmusbrandt/FastConvergentCoordinatedPrecoding.jl.

APPENDIX A
PROOF OF UNIQUENESS FOR $\mathbf{U}_{i_k}^\star$

First note that $\boldsymbol{\Phi}_{i_k}, \mathbf{Z}_{i_k}, \mathbf{Y}_{i_k}$ all are positive definite matrices and that $\boldsymbol{\Psi}_{i_k}$ is positive semidefinite. The necessary and sufficient optimality condition for $\mathbf{U}_{i_k}$ in (6) can be written as

$$\rho \boldsymbol{\Psi}_{i_k} \mathbf{U}_{i_k}^\star \mathbf{Z}_{i_k} + \boldsymbol{\Phi}_{i_k} \mathbf{U}_{i_k}^\star \mathbf{Y}_{i_k} = \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \mathbf{Y}_{i_k}$$

which in its vectorized form is

$$\left(\rho \left(\mathbf{Z}_{i_k}^\mathsf{T} \otimes \boldsymbol{\Psi}_{i_k}\right) + \left(\mathbf{Y}_{i_k}^\mathsf{T} \otimes \boldsymbol{\Phi}_{i_k}\right)\right) \operatorname{vec}\left(\mathbf{U}_{i_k}^\star\right) = \operatorname{vec}\left(\mathbf{H}_{i_k i} \mathbf{V}_{i_k} \mathbf{Y}_{i_k}\right). \tag{8}$$

Now,

$$\begin{aligned}
&\lambda_{\text{min}}\left(\rho\left(\mathbf{Z}_{i_k}^\mathsf{T} \otimes \boldsymbol{\Psi}_{i_k}\right) + \left(\mathbf{Y}_{i_k}^\mathsf{T} \otimes \boldsymbol{\Phi}_{i_k}\right)\right) \\
&\overset{\text{(a)}}{\geq} \rho\, \lambda_{\text{min}}\left(\mathbf{Z}_{i_k}^\mathsf{T} \otimes \boldsymbol{\Psi}_{i_k}\right) + \lambda_{\text{min}}\left(\mathbf{Y}_{i_k}^\mathsf{T} \otimes \boldsymbol{\Phi}_{i_k}\right) \\
&\overset{\text{(b)}}{=} \rho\, \lambda_{\text{min}}\left(\mathbf{Z}_{i_k}\right) \lambda_{\text{min}}\left(\boldsymbol{\Psi}_{i_k}\right) + \lambda_{\text{min}}\left(\mathbf{Y}_{i_k}\right) \lambda_{\text{min}}\left(\boldsymbol{\Phi}_{i_k}\right) \\
&\overset{\text{(c)}}{\geq} \lambda_{\text{min}}\left(\mathbf{Y}_{i_k}\right) \lambda_{\text{min}}\left(\boldsymbol{\Phi}_{i_k}\right) \\
&\overset{\text{(d)}}{>} 0
\end{aligned}$$

where (a) follows from Weyl's inequality and the Hermitianness of a Kronecker product of two Hermitian matrices, (b) follows from the Kronecker structure, and (c) and (d) follows from the positive (semi-)definiteness of the involved matrices. Thus, the square coefficient matrix in (8) is non-singular, and the system of equations has a unique solution $\mathbf{U}_{i_k}^\star$.

APPENDIX B
DETAILS OF BISECTION FOR $\mu_i^\star$

The optimal $\mu_i^\star$ can be found efficiently by introducing the eigenvalue decompositions $\boldsymbol{\Gamma}_i + \rho \boldsymbol{\Lambda}_{i_k} = \mathbf{L}_{i_k} \boldsymbol{\Delta}_{i_k} \mathbf{L}_{i_k}^\mathsf{H}$ and rewriting the sum power $f_i(\mu_i) \triangleq \sum_{k=1}^{K_i} \left\|\mathbf{V}_{i_k}^\star\right\|_\mathsf{F}^2$ as

$$f_i(\mu_i) = \sum_{k=1}^{K_i} \alpha_{i_k}^2 \sum_{m=1}^{M_i} \frac{\left[\mathbf{L}_{i_k}^\mathsf{H} \mathbf{H}_{i_k i}^\mathsf{H} \mathbf{U}_{i_k} \mathbf{Y}_{i_k} \mathbf{Y}_{i_k}^\mathsf{H} \mathbf{U}_{i_k}^\mathsf{H} \mathbf{H}_{i_k i} \mathbf{L}_{i_k}\right]_{mm}}{\left(\left[\boldsymbol{\Delta}_{i_k}\right]_{mm} + \mu_i\right)^2}.$$

The optimal $\mu_i^\star$ can then be found by bisection of $f_i(\mu_i)$ on $(0, \mu_i^{\text{max}}]$ where

$$\mu_i^{\text{max}} = \sqrt{\frac{1}{P_i} \cdot \max_{k,m}\, c_{i_k,m} \cdot M_i \cdot \sum_{k=1}^{K_i} \alpha_{i_k}^2} - \min_{k,m}\, \left[\boldsymbol{\Delta}_{i_k}\right]_{mm},$$

and $c_{i_k,m} = \left[\mathbf{L}_{i_k}^\mathsf{H} \mathbf{H}_{i_k i}^\mathsf{H} \mathbf{U}_{i_k} \mathbf{Y}_{i_k} \mathbf{Y}_{i_k}^\mathsf{H} \mathbf{U}_{i_k}^\mathsf{H} \mathbf{H}_{i_k i} \mathbf{L}_{i_k}\right]_{mm}$.

APPENDIX C
ALGORITHM RUN TIMES

As an indication of algorithm complexity, we provide a simple benchmark of the wall-clock run times of the different algorithms. The algorithms were implemented in the Julia language [14], and are available for download at [13]. The reweighted RCRM is based on semidefinite programming; for this we use Convex.jl [15] together with the Mosek solver (version 7.1.0.32, single-threaded) [16]. We run a benchmark where the algorithms are run on $L_{\text{bench}} = 10$ realizations of the $(3 \times 2, 1)^6$ interference channel at $\texttt{SNR} = 30\,\text{dB}$. We let $L_{\text{OTA}} = 3$ for all algorithms and $L_{\text{local}} = 4$ for FastDCP. The benchmark was run on a 2011 Apple MacBook Pro with a quad-core 2 GHz Intel Core i7 processor and 8 GB of RAM. The results can be seen below in Table I. The fact that FastDCP and WMMSE are based on closed-form solutions are immediately obvious when comparing to the run time of reweighted RCRM.

TABLE I. WALL-CLOCK RUN TIMES

Algorithm	Run Time [s]
FastDCP	0.109
WMMSE [7]	0.0334
Reweighted RCRM [11]	9.77

APPENDIX REFERENCES

[14] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *arXiv:1411.1607 [cs.MS]*, 2014.

[15] M. Udell, K. Mohan, D. Zeng, J. Hong, S. Diamond, and S. Boyd, "Convex optimization in Julia," in *Proc. Workshop High Performance Tech. Comput. Dynamic Languages (HPTCDL'14)*, 2014.

[16] MOSEK, "The MOSEK C optimizer API manual, v7.1," 2014.