A Successive Optimization Approach to Pilot Design for Multi-Cell Massive MIMO Systems

Hayder Al-Salihi, Trinh Van Chien, Tuan Anh Le, and Mohammad Reza Nakhai

Abstract—In this letter, we introduce a novel pilot design approach that minimizes the total mean square errors of the minimum mean square error estimators of all base stations (BSs) subject to the transmit power constraints of individual users in the network, while tackling the pilot contamination in multi-cell Massive MIMO systems. First, we decompose the original non-convex problem into distributed optimization sub-problems at individual BSs, where each BS can optimize its own pilot signals given the knowledge of pilot signals from the remaining BSs. We then introduce a successive optimization approach to transform each optimization sub-problem into a linear matrix inequality (LMI) form, which is convex and can be solved by available optimization packages. Simulation results confirm the fast convergence of the proposed approach and prevails a benchmark scheme in terms of providing higher accuracy.

I. INTRODUCTION

In multi-cell Massive MIMO systems, each base station (BS) requires accurate knowledge of the channel state information (CSI) obtained during the pilot training phase. To attain accurate channel estimates, perfectly orthogonal pilot allocations to users are required. Unfortunately, this requirement is impractical, since the pilot overhead has to be proportional to the number of users in the entire system. Furthermore, the channel coherence block limits the number of orthogonal pilots [1]. Thus, pilot signals need to be reused over cells, causing spatially correlated interference, known as pilot contamination that degrades the performance of a Massive MIMO system [1].

In order to address the pilot contamination problem, the authors of [2] proposed a superimposed channel estimation approach by adding a low power pilot signal to the data signal at the transmitter. The superimposed signal is then utilized at the receiver for channel estimation. However, a proportion of the power allocated to the pilot signal is wasted. Fortunately, it has been shown in [3] that the wasted-power problem can be theoretically mitigated with properly designed forward-error-correction codes. On the other attempts, pilot assignment and pilot power control are alternative solutions which can attain great improvements for the case that the system only has a finite set of orthogonal pilot signals [4]–[6]. Involving reuse factor in pilot design may lead to a combinatorial pilot

assignment problem in many pilot designs and, hence, result in an exponentially increased computational complexity [7].

In this letter, we consider a multi-cell Massive MIMO system adopting minimum mean square error (MMSE) estimators at BSs. We derive the mean square error (MSE) of the adopted MMSE estimator as a widely used accuracy criteria for estimation. We, then, formulate an optimization problem to find optimal pilot signals that minimize the total derived MSE of the MMSE estimators of all BSs in the network subject to a transmit power constraint at each user. The proposed formulation is non-convex with respect to the pilot matrices. To overcome non-convexity, we, first, decompose the proposed optimization problem into distributed subproblems at BSs, where each BS in the network optimizes its own pilot signal, given the knowledge of the pilot signals of other BSs. We then introduce a successive optimization approach to transform each subproblem into a linear matrix inequality (LMI) problem which is convex and can be effectively solved by available optimization packages, e.g., CVX [8]. Finally, we analyse the complexity of the transformed LMI optimization problem.

Notation: Bold lower/upper case letters are used for vectors/matrices; $\|\cdot\|_F$ and $\|\cdot\|$ stand for the Frobenius norm and the Euclidean norm; $(\cdot)^T$ and $(\cdot)^H$ is the regular and complex conjugate transpose operator, respectively; $\mathrm{Tr}(\cdot)$ is the trace of a matrix; $\mathbf{X} \succeq \mathbf{0}$ is the positive semidefinite condition; $\mathbf{I}_a$ is an $a \times a$ identity matrix; $\mathrm{diag}\{\mathbf{x}\}$ is a diagonal matrix which the diagonal entries are elements of the vector $\mathbf{x}$; $\mathcal{CN}(\cdot, \cdot)$ is a circularly symmetric complex Gaussian distribution; $\mathbb{E}[\cdot]$ is the expectation of a random variable; $\mathcal{O}(\cdot)$ is the big-O notation.

II. SYSTEM MODEL

Consider a multi-cell massive MIMO system with C cells operating in a time-division duplexing mode. Each cell comprises of an M-antenna BS and N single-antenna users. The propagation factor between the i-th antenna of the BS in cell c^*, denoted as BS c^*, and user n in cell c is $\sqrt{\phi_{c,c^*}^n} h_{c,c^*,i}^n$, where ϕ_{c,c^*}^n is the large scale fading coefficient modeling the path-loss and shadowing, while $h_{c,c^*,i}^n \sim \mathcal{CN}(0,1)$ is small-scale fading.

In the pilot training phase, all users in each cell synchronously send their pilot signals. Let $\mathbf{x}_c^n \in \mathbb{C}^{\tau \times 1}$ be the pilot signal used by user n in cell c and $\|\mathbf{x}_c^n\|^2 \leq P_{\max,c}, \forall c$, where τ is the length of the pilot signal, and $P_{\max,c}$ is the maximum allocated power level by each user in cell c to its pilot signal. The received baseband training signal $\mathbf{y}_{c^*,i} \in \mathbb{C}^{\tau \times 1}$ at the i-th antenna element of the BS c^* can be expressed as:

$$\mathbf{y}_{c^*,i} = \sum_{c=1}^{C} \sum_{n=1}^{N} \sqrt{\phi_{c,c^*}^n} h_{c,c^*,i}^n \mathbf{x}_c^n + \mathbf{v}_{c^*,i}, \quad (1)$$

where $\mathbf{v}_{c^*,i}$ is Gaussian noise with $\mathbf{v}_{c^*,i} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_\tau)$. Let the received signals, Gaussian noises, pilot signals by all antenna elements of BS c^* and the corresponding large scale channel coefficients be denoted as

$$\mathbf{Y}_{c^*} = [\mathbf{y}_{c^*,1},\ \mathbf{y}_{c^*,2}, \ldots, \mathbf{y}_{c^*,M}] \in \mathbb{C}^{\tau \times M}, \tag{2}$$

$$\mathbf{V}_{c^*} = [\mathbf{v}_{c^*,1},\ \mathbf{v}_{c^*,2}, \ldots,\ \mathbf{v}_{c^*,M}] \in \mathbb{C}^{\tau \times M}, \tag{3}$$

$$\mathbf{X}_{c^*} = [\mathbf{x}_{c^*}^1,\ \mathbf{x}_{c^*}^2, \ldots, \mathbf{x}_{c^*}^N] \in \mathbb{C}^{\tau \times N}, \tag{4}$$

$$\mathbf{D}_{c,c^*} = \text{diag}\{[\phi_{c,c^*}^1,\ \phi_{c,c^*}^2, \ldots, \phi_{c,c^*}^N]^T\} \in \mathbb{C}^{N \times N}. \tag{5}$$

Also, let the small-scale fading channel coefficients of all N users in cell c as seen by BS c^* be expressed as

$$\mathbf{H}_{c,c^*} = \begin{bmatrix} h_{c,c^*,1}^1, & \cdots, & h_{c,c^*,M}^1 \\ \vdots & \ddots & \vdots \\ h_{c,c^*,1}^N & \cdots & , h_{c,c^*,M}^N \end{bmatrix} \in \mathbb{C}^{N \times M}. \tag{6}$$

Then, using $(1) - (6)$, one can formulate the received training signals by all M antenna elements of BS c^*, according

$$\mathbf{Y}_{c^*} = \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{H}_{c^*,c^*} + \sum_{c=1, c \neq c^*}^{C} \mathbf{X}_c \mathbf{D}_{c,c^*}^{\frac{1}{2}} \mathbf{H}_{c,c^*} + \mathbf{V}_{c^*}. \tag{7}$$

The first term in right hand side of (7) involves desired channel coefficients and the remaining terms indicate the effects of mutual interference and Gaussian noise. The channel estimate $\widehat{\mathbf{H}}_{c^*,c^*}$ of the original channel $\mathbf{H}_{c^*,c^*}$ is computed by utilizing the MMSE estimation upon the observation of $\mathbf{Y}_{c^*}$ is [9]:

$$\widehat{\mathbf{H}}_{c^*,c^*} = \mathbb{E}[\mathbf{H}_{c^*,c^*} \mathbf{Y}_{c^*}^H] \left(\mathbb{E}[\mathbf{Y}_{c^*} \mathbf{Y}_{c^*}^H] \right)^{-1} \mathbf{Y}_{c^*}. \tag{8}$$

Plugging (7) in (8), and after some mathematical manipulations, we obtain

$$\widehat{\mathbf{H}}_{c^*,c^*} = M \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{\Omega}_{c^*}^{-1} \mathbf{Y}_{c^*}, \tag{9}$$

where $\mathbf{\Omega}_{c^*} = M \sum_{c=1}^{C} \mathbf{X}_c \mathbf{D}_{c,c^*} \mathbf{X}_c^H + M \sigma^2 \mathbf{I}_\tau$. From (9), the channel estimation quality depends on the pilot design and if $\tau < CN$, it also suffers from pilot contamination [4]–[6]. Let the channel estimation errors at BS c^* be denoted as

$$\mathbf{\Delta}_{c^*} = \mathbf{H}_{c^*,c^*} - \widehat{\mathbf{H}}_{c^*,c^*}, \tag{10}$$

and the MSE be defined as

$$\text{MSE}_{c^*} = \mathbb{E}\left[\|\mathbf{\Delta}_{c^*}\|_F^2 \right] = \mathbb{E}\left[\text{Tr}\left(\mathbf{\Delta}_{c^*} \mathbf{\Delta}_{c^*}^H \right) \right]. \tag{11}$$

Then, using (7), (8), (9), and after some mathematical manipulations, one can rewrite MSE in (11) as:

$$M\text{Tr}\left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1} \right), \tag{12}$$

where $\mathbf{A}^{-1} = \mathbf{I}_N$, $\mathbf{B} = M \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H$, $\mathbf{D} = \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}$, and $\mathbf{C}^{-1} = M \sum_{c=1, c \neq c^*}^{C} \mathbf{X}_c \mathbf{D}_{c,c^*} \mathbf{X}_c^H + M \sigma^2 \mathbf{I}_\tau$. By utilizing the Sherman-Morrison-Woodbury identity

$$(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}$$

and defining $\text{MSE}_{c^*} \triangleq f_{c^*}(\mathbf{X}_{c^*})$, one can reformulate (11) as

$$f_{c^*}(\mathbf{X}_{c^*}) = M\text{Tr}\left(\left(\mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{F}_{c^*}^{-1} \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \right)^{-1} \right), \tag{13}$$

where $\mathbf{F}_{c^*} = \sum_{c=1, c \neq c^*}^{C} \mathbf{X}_c \mathbf{D}_{c,c^*} \mathbf{X}_c^H + \sigma^2 \mathbf{I}_\tau$.

III. A Successive Optimization Pilot Design

From (13), the performance of the MMSE estimation algorithm depends on the pilot structure. In this section, we develop an optimal pilot design to minimize the total channel estimation errors of all BSs in the network subject to the transmit power constraints at individual users. Hence, we introduce the following optimization problem for the network:

$$\begin{aligned} \underset{\{\mathbf{X}_{c^*}\}}{\text{minimize}} \quad & \sum_{c^*=1}^{C} f_{c^*}(\mathbf{X}_{c^*}) \\ \text{subject to} \quad & \mathbf{X}_{c^*}^H \mathbf{X}_{c^*} \preceq P_{\max,c^*} \mathbf{I}_N, \forall c^*, \end{aligned} \tag{14}$$

where $\{\mathbf{X}_{c^*}\} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_C\}$. Problem (14) is non-convex due to its objective function. To tackle the problem, we first introduce an auxiliary variable $\mathbf{G}_{c^*}$, denote $\{\mathbf{G}_{c^*}\} = \{\mathbf{G}_1, \cdots, \mathbf{G}_C\}$, remove the constant M, and rewrite (14) as[1]

$$\begin{aligned} \underset{\{\mathbf{X}_{c^*}\},\{\mathbf{G}_{c^*}\}}{\text{minimize}} \quad & \sum_{c^*=1}^{C} \text{Tr}(\mathbf{G}_{c^*}) \\ \text{subject to} \quad & \mathbf{X}_{c^*}^H \mathbf{I}_\tau^{-1} \mathbf{X}_{c^*} \preceq P_{\max,c^*} \mathbf{I}_N, \forall c^*, \\ & \left(\mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{F}_{c^*}^{-1} \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \right)^{-1} \preceq \mathbf{G}_{c^*}, \forall c^*. \end{aligned} \tag{15}$$

Adopting the Schur complement [10], one can equivalently reformulate problem (15) as

$$\begin{aligned} \underset{\{\mathbf{X}_{c^*}\},\{\mathbf{G}_{c^*}\}}{\text{minimize}} \quad & \sum_{c^*=1}^{C} \text{Tr}(\mathbf{G}_{c^*}) \\ \text{subject to} \quad & \begin{bmatrix} P_{\max,c^*}\mathbf{I}_N & \mathbf{X}_{c^*}^H \\ \mathbf{X}_{c^*} & \mathbf{I}_\tau \end{bmatrix} \succeq \mathbf{0}, \forall c^*, \\ & \begin{bmatrix} \mathbf{G}_{c^*} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{F}_{c^*}^{-1} \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \end{bmatrix} \succeq \mathbf{0}, \forall c^*. \end{aligned} \tag{16}$$

The second set of constraints in (16) is still non-convex due to the nonlinearity of the term $\mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{F}_{c^*}^{-1} \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}$ with respect to optimization variable $\mathbf{X}_{c^*}, \forall c^*$, i.e., the optimization variable is in quadratic forms and appears in both numerator and denominator of the term. As a main contribution of this paper, we propose a distributed algorithm where every BS c^* optimizes its own pilot signals given the knowledge of the pilot signals of the other cells in $\mathbf{F}_{c^*}^{-1}$ as follows:

$$\begin{aligned} \underset{\mathbf{X}_{c^*},\mathbf{G}_{c^*}}{\text{minimize}} \quad & \text{Tr}(\mathbf{G}_{c^*}) \\ \text{subject to} \quad & \begin{bmatrix} P_{\max,c^*}\mathbf{I}_N & \mathbf{X}_{c^*}^H \\ \mathbf{X}_{c^*} & \mathbf{I}_\tau \end{bmatrix} \succeq \mathbf{0}, \\ & \begin{bmatrix} \mathbf{G}_{c^*} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{F}_{c^*}^{-1} \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{17}$$

Although the distributed optimization problem (17) only considers $\mathbf{X}_{c^*}$ and $\mathbf{G}_{c^*}$ as the optimization variables, its second constraint is still not in an LMI form with respect to $\mathbf{X}_{c^*}$. To proceed, we propose a successive optimization approach

[1]Problems (14) and (15) are equivalent since (15) is an epigraph form of (14) [10, pp.134]. In fact, introducing the auxiliary variable $\mathbf{G}_{c^*}$ transforms the objective function into a linear form while shifting the nonlinear part into a constraint.

where, at the t-th iteration, BS c^* updates its pilot signals by solving the following distributed optimization problem:

$$\underset{\mathbf{X}_{c^*}^{(t)}, \mathbf{G}_{c^*}^{(t)}}{\text{minimize}} \quad \mathrm{Tr}\left(\mathbf{G}_{c^*}^{(t)}\right)$$

$$\text{subject to} \quad \begin{bmatrix} P_{\max,c^*}\mathbf{I}_N & \mathbf{X}_{c^*}^{(t),H} \\ \mathbf{X}_{c^*}^{(t)} & \mathbf{I}_\tau \end{bmatrix} \succeq \mathbf{0},$$

$$\begin{bmatrix} \mathbf{G}_{c^*}^{(t)} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\mathbf{X}_{c^*}^{(t),H}(\mathbf{F}_{c^*}^{-1})^{(t-1)}\mathbf{X}_{c^*}^{(t-1)}\mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \end{bmatrix} \succeq \mathbf{0},$$
(18)

where $\mathbf{F}_{c^*}^{-1}$ from the previous iteration is

$$(\mathbf{F}_{c^*}^{-1})^{(t-1)} = \sum_{c=1,c\neq c^*}^{C} \mathbf{X}_c^{(t-1)}\mathbf{D}_{c,c^*}\mathbf{X}_c^{(t-1),H} + \sigma^2\mathbf{I}_\tau, \quad (19)$$

$\mathbf{X}_{c^*}^{(t-1)}$ and $\mathbf{X}_c^{(t-1)}$ are the optimal pilots of cells c^* and c, respectively, which are obtained from the $(t-1)$-th iteration. In order to transform the second constraint of (18) into an LMI form with respect to both $\mathbf{X}_{c^*}^{(t)}$ and $\mathbf{G}_{c^*}^{(t)}$, we have used the known value of $\mathbf{X}_{c^*}^{(t-1)}$. Notice that at the stationary point attained after a sufficient number of iterations, the approximation

$$\mathbf{X}_{c^*}^{(t)} \approx \mathbf{X}_{c^*}^{(t-1)}, \forall c^*, \quad (20)$$

can be assured with any desired accuracy. Note that, the matrix in the second constraint of (18) is not Hermitian during the iterations, due to the mismatch between $\mathbf{X}_{c^*}^{(t)}$ and $\mathbf{X}_{c^*}^{(t-1)}, \forall c^*$. To guarantee a Hermitian matrix in the second constraint of (18), we introduce a new variable $\mathbf{A}_{c^*}^{(t)}$, such that

$$2\mathbf{A}_{c^*}^{(t)} = \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\mathbf{X}_{c^*}^{(t),H}(\mathbf{F}_{c^*}^{-1})^{(t-1)}\mathbf{X}_{c^*}^{(t-1)}\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}$$
$$+ \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\mathbf{X}_{c^*}^{(t-1),H}(\mathbf{F}_{c^*}^{-1})^{(t-1)}\mathbf{X}_{c^*}^{(t)}\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}. \quad (21)$$

Finally, we reformulate (18) as

$$\underset{\mathbf{X}_{c^*}^{(t)}, \mathbf{G}_{c^*}^{(t)}, \mathbf{A}_{c^*}^{(t)}}{\text{minimize}} \quad \mathrm{Tr}\left(\mathbf{G}_{c^*}^{(t)}\right)$$

$$\text{subject to} \quad \begin{bmatrix} P_{\max,c^*}\mathbf{I}_N & \mathbf{X}_{c^*}^{(t),H} \\ \mathbf{X}_{c^*}^{(t)} & \mathbf{I}_\tau \end{bmatrix} \succeq \mathbf{0},$$

$$\begin{bmatrix} \mathbf{G}_{c^*}^{(t)} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{A}_{c^*}^{(t)} \end{bmatrix} \succeq \mathbf{0},$$
(22)

$$\text{constraint } (21).$$

Problem (22) is now convex and can be efficiently solved by CVX [8]. The procedure to obtain the optimal pilot signals for all C cells in the network is summarized in Algorithm 1.

Remark 1 (Convergence): Since problem (22) is convex, steps 3 and 4 in Algorithm 1 ensure the δ-convergence of $\mathbf{X}_{c^*}^{(t)}$ to its optimal value and a minimal objective function value in problem (22) per cell.[2]

As the main computational complexity of Algorithm 1 is to solve (22) at each BS, we analyze such complexity in the sequel. Since (22) contains LMI constraints, a standard

---

[2]Although the global optimality can be achieved per iteration and per cell by solving (22), it may not be achievable to the original multicell problem (14) due to its inherent non-convexity. In fact, Algorithm 1 is a suboptimal algorithm with an affordable complexity to the NP-hard problem (14).

Algorithm 1 Successive optimization approach for (15)

1: **Inputs:** $\mathbf{D}_{c,c^*}$, $P_{\max,c^*}$, σ^2, stopping criteria $\delta > 0$, initialize $\mathbf{X}_c^{(0)}$, $\forall c, c^*$; $t = 1$;
2: Each cell c^* calculates $\mathbf{F}_{c^*}^{(t-1)}$ utilizing (19) and then solves (22) to attain $\mathbf{X}_{c^*}^{(t)}$, $\forall c^*$; Exchange $\mathbf{X}_{c^*}^{(t)}$ with the other cells;
3: **If** $\sum_{c^*=1}^{C}\|\mathbf{X}_{c^*}^{(t)} - \mathbf{X}_{c^*}^{(t-1)}\|_F \leq \delta$, **then** Go to step 5;
4: **else if** $\sum_{c^*=1}^{C}\|\mathbf{X}_{c^*}^{(t)} - \mathbf{X}_{c^*}^{(t-1)}\|_F > \delta$, **then** $t = t+1$; Go to step 2;
5: **Outputs:** $\mathbf{X}_{c^*}^{\star} \leftarrow \mathbf{X}_{c^*}^{(t)}$, $\forall c^*$.

---

interior-point method (IPM) [10] can be used to find its optimal solution. Therefore, we consider the worst-case runtime of the IPM to analyze the computational complexities of the proposed problem (22) as follows.

Definition 1: For a given $\epsilon > 0$, the set of $\mathbf{X}_{c^*}^{(t),\epsilon}, \mathbf{G}_{c^*}^{(t),\epsilon}, \mathbf{A}_{c^*}^{(t),\epsilon}$ is called an ϵ-solution to (22) if

$$\mathrm{Tr}\left(\mathbf{G}_{c^*}^{(t),\epsilon}\right) \leq \underset{\mathbf{X}_{c^*}^{(t)}, \mathbf{G}_{c^*}^{(t)}, \mathbf{A}_{c^*}^{(t)}}{\text{minimize}} \mathrm{Tr}\left(\mathbf{G}_{c^*}^{(t)}\right) + \epsilon. \quad (23)$$

It can be observed that the number of decision variables of problem (22) is on the order of $(2N + \tau)N$. Let $m = \mathcal{O}\left((2N + \tau)N\right)$, we introduce the following lemma.

Lemma 1: The computational complexities to obtain ϵ-solution to problem (22) is

$$\ln(\epsilon^{-1})\sqrt{4N + \tau}\,\alpha m, \quad (24)$$

where $\alpha = 10N^3 + (3\tau + 6m)N^2 + Nm\tau(m\tau + 2) + \tau^2(m + \tau) + m^2$.

Proof: Problem (22) has 1 LMI constraint of dimension $N + \tau$, 1 LMI constraint of dimension $2N$, and 1 LMI constraint of dimension N. Based on these observations, one can follow the same steps as in [11, Section V-A] to arrive at (24). Note that the term $\ln(\epsilon^{-1})\sqrt{4N + \tau}$ in (24) is the iteration complexities [11] required for obtaining ϵ-solutions to problem (22) while the remaining terms represent the per-iteration computation costs [11]. ∎

IV. SIMULATION RESULTS

A wrapped-around multi-cell Massive MIMO system is considered for simulations with $C = 4$, $M = 500$, and $N = 10$. All users are randomly distributed over the coverage area. However, the distance between any user n of cell c and BS c^*, denoted as d_{c,c^*}^n, is always satisfied $d_{c,c^*}^n \geq 0.035$ km. The system utilizes 20 MHz bandwidth related to the noise variance of -96 dBm and the noise figure of 5 dB. The large-scale fading coefficient ϕ_{c,c^*}^n [dB] is modeled as $\phi_{c,c^*}^n = -148.1 - 37.6\log_{10}(d_{c,c^*}^n) + z_{c,c^*}^n$, where z_{c,c^*}^n is the shadow fading following a log-normal Gaussian distribution with the standard variation of 7 dB. Monte-Carlo simulations are tackled over 200 different realizations of user locations. The widely adopted orthogonal pilot design, e.g., [4], [12], is used as a benchmark where each pilot symbol is allocated with power 200 mW and those orthogonal pilots are reused amongst users in the network. For every realization of user locations, such pilot signals are generated by the eigenvectors
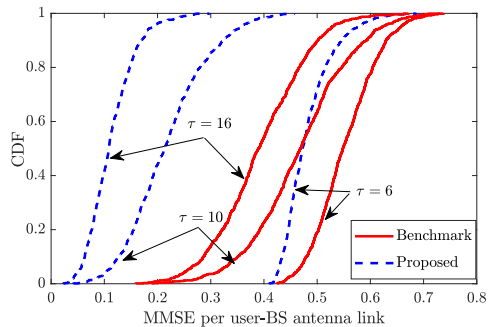
Figure 1. The CDFs of the MMSE per user-BS antenna link for the proposed and benchmark approaches.
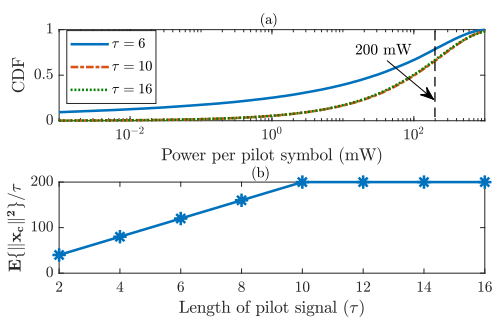

Figure 2. (*a*) The CDFs of the power allocated for each pilot symbol of the proposed approach with different pilot lengths; (*b*) The average power allocated for each pilot symbol of the proposed approach v.s. the pilot length.

of a uniformly generated random matrix. The power constraint for pilot signal is set to be $P_{\max,c} = 200\tau$ mW, $\forall c$.

Fig. 1 shows the cumulative distribution function (CDF) of the BS-user antenna link which is defined by $f_{c^*}(\mathbf{X}_{c^*})/(MN)$. It is clear from the figure that the channel estimation accuracy of the proposed approach is significantly improved compared to that of the benchmark. This confirms the effectiveness of our optimal pilot design in combating pilot contaminations. The results also indicate that the performance gap between the proposed approach and the benchmark increases as the pilot length increases. This is because increasing pilot length gives more degrees of freedom to the proposed approach for optimizing its performances.[3]

Fig. 2 (a) displays the CDFs of the power allocated to each pilot symbol with different pilot lengths. It can be seen from the figure that for most of the cases, e.g., around 80 % for $\tau = 6$ and 70 % for $\tau = 10$ and $\tau = 16$, the proposed approach spends less power for each symbol than the benchmark does, i.e., less than 200 mW per symbol. Fig. 2 (b) illustrates the average power for each pilot symbol against the length of pilot signals. When $\tau < N = 10$, the proposed approach consumes less power than the benchmark does and the system suffers from both intra-cell and inter-cell interference. In such hostile situations, the proposed approach can still effectively handle the pilot contaminations which results in higher channel

---

[3]The MMSE of the system approaches zero when the pilot length goes to infinity. An ideal pilot length of $\tau = NC$ is sufficient to distinguish all users in the network, and also to balance between channel estimation errors and spectral efficiency (SE). However, this is impractical for a large-scale network. To that end, the proposed approach offers significant MMSE improvements for practical pilot lengths, i.e., when $\tau < NC$.
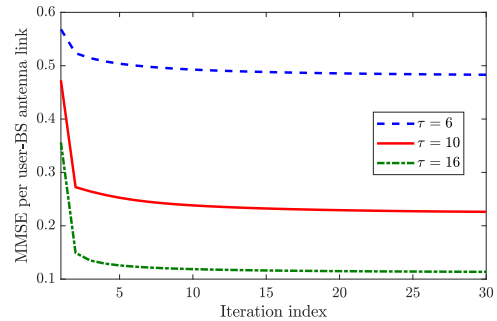

Figure 3. The convergence of MMSE per user-BS antenna link v.s. iteration index for the proposed approach.

Table I
AVERAGE UPLINK SE OF EACH USER IN THE NETWORK [BITS/S/HZ] FOR DIFFERENT PILOT LENGTHS.

	$\tau = 6$	$\tau = 10$	$\tau = 16$
Benchmark	1.19	1.49	1.55
Proposed	1.39	1.74	1.86
Gain	16.81%	16.78%	20.00%

estimate accuracy as seen in Fig. 1, while consuming less power than the benchmark. Interestingly, when $\tau \geq N$, the optimal transmit power spent on each symbol turns out to be 200 mW and to be constant irrespective to the value of τ, which can also be observed from the CDF shown in Fig. 2 (a).

Fig. 3 numerically reveals the fast convergent speed of the proposed approach, i.e., within less than 20 iterations. This results confirm the statement in Remark 1. Finally, Table I demonstrates the average uplink SE of per user in the network using the use-and-then-forget capacity bounding technique [6, eq. (28)] with maximum ratio detection, fixed data power 200 mW, and the coherence block length 200 symbols. Thanks to minimizing MSE, the proposed approach attains a higher uplink SE than the benchmark does.

REFERENCES

[1] E. Björnson *et al.*, "Massive MIMO: 10 Myths and One Critical Question",vol. 54, no. 2, pp. 114-123, 2016.
[2] H. Zhang *et al.*, "On superimposed pilot for channel estimation in multicell multiuser mimo uplink: Large system analysis," *IEEE Trans. Vehi. Techno.*, vol. 65, no. 3, pp. 1492–1505, 2016.
[3] J. Ma *et al.*, "On Orthogonal and Superimposed Pilot Schemes in Massive MIMO NOMA Systems," *IEEE Journal on Selected Areas in Commun.*, vol. 35, no. 12, pp. 2696-2707, 2017.
[4] X. Zhu *et al.*, "Smart pilot assignment for Massive MIMO," *IEEE Commun. Letters*, vol. 19, no. 9, pp. 1644 - 1647, 2015.
[5] P. Wang *et al.*, "A Novel Pilot Assignment Approach for Pilot Decontaminating in Massive MIMO Systems, " in *Proc.* IEEE Wireless Commun. and Networking Conf. (WCNC), pp. 1-6, 2017.
[6] T. V. Chien *et al.*, "Joint Pilot Design and Uplink Power Allocation in Multi-Cell Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, early access.
[7] S. Ma, *et al.*, "A Novel Pilot Assignment Scheme in Massive MIMO Networks," *IEEE Wireless Commun. Letters*, early access.
[8] M. C. Grant and S. P. Boyd, *The CVX Users' Guide, Release 2.1.*, Mar. 2015, [Online]. Available: http://web.cvxr.com/cvx/doc/CVX.pdf.
[9] T. Kailath *et al.*, *Linear Estimation*. Prentice Hall, 2000.
[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[11] K.-Y. Wang *et al.*, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690-5705, 2014.
[12] L. You *et al.*, "Pilot reuse for Massive MIMO transmission over spatially correlated rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352-3366, 2015.