Pareto-Optimal Pilot Design for Cellular Massive MIMO Systems

Tuan Anh Le, *Senior Member IEEE*, Trinh Van Chien, Mohammad Reza Nakhai, *Senior Member IEEE*, and Tho Le-Ngoc, *Fellow IEEE*

***Abstract*—We introduce a non-orthogonal pilot design scheme that simultaneously minimizes two contradicting targets of channel estimation errors of all base stations (BSs) and the total pilot power consumptions of all users in a multi-cell massive MIMO system, subject to the transmit power constraints of the users in the network. We formulate a multi-objective optimization problem (MOP) with two objective functions capturing the contradicting targets and find the Pareto optimal solutions for the pilot signals. Using weighted-sum-scalarization technique, we first convert the MOP to an equivalent single-objective optimization problem (SOP), which is not convex. Assuming that each BS is provided with the most recent knowledge of the pilot signals of the other BSs, we then decompose the SOP into a set of distributed non-convex optimization problems to be solved at individual BSs. Finally, we introduce an alternating optimization approach to cast each one of the resulting distributed optimization problems into a convex linear matrix inequality (LMI) form. We provide a mathematical proof for the convergence of the proposed alternating approach and a complexity analysis for the LMI optimization problem. Simulation results confirm that the proposed approach significantly reduces pilot power, whilst maintaining the same level of channel estimation error as in [1].**

***Index Terms*—Massive MIMO, Pilot Design, Linear Matrix Inequality, Multi-objective Optimization**

I. Introduction

In order to fully realize the potential of massive MIMO networks, pilot contamination needs to be properly addressed [2]. Pilot contamination can be tackled by wise pilot assignments exploiting multiple domains. The same pilot sequence is assigned to different users having non-overlapping angle-of-arrival (AoA) at the corresponding BSs [3], [4], or distinguishable power profiles, i.e., power-delay profile, the angular power spectra and the Doppler power spectra [5], [6]. Pointing out the fact that most of those schemes assign pilot sequences to users without considering the severity of pilot contamination being different from user to user, the authors of [7] formulate their pilot assignment problem as a minimum-weight multi-index assignment problem guaranteeing the fairness of all users in the network. Although pilot assignment approaches exploit different dimensions to maintain the orthogonality between pilot sequences allocated to users within the network, their performances are still limited by a certain amount of users being served. It is also noted that the power spent on the pilot sequence of every user is fixed.

The other trend is to accept the non-orthogonality of pilot sequences assigned amongst users in the network while controlling the power allocated to those pilot sequences to eliminate the pilot contamination problem. Pilot power allocation problems are normaly cast as optimization problems where the optimization variables are pilot powers. To this end, maximizing the mutual information between the channel measurements and the actual channel vector was introduced in [8]. Maximizing the system spectral efficiency was considered in [9] while minimizing the channel estimation mean square error (MSE) was addressed in [1], [10]–[13]. Pilot power can also be jointly designed with the downlink power, i.e., the power spent to convey downlink data. For example, the joint design in [14] guarantees that the user capacity of the network is attained and the pre-defined SINR levels of all users are met while the consideration in [15] is the max-min fairness problem under power budget constraints. Although the aforementioned designs can either improve the accuracy of the channel estimators, e.g., [1], or the spectral efficiency, e.g., [9], power efficiency is not the focus of those methods. On the other hand, power efficiency design is one of the key factors leading to the energy efficiency [16], [17] identified as a major concern for the deployment and operation of a wireless communication network [16], [18].

In this paper, we aim to improve both channel estimation accuracy and power efficiency. Particularly, we consider a multi-cell massive MIMO system, where minimum mean square error (MMSE) estimators are employed at BSs, and focus on minimizing two objectives: i) the total channel estimation errors at BSs; ii) the total pilot power consumed by all mobile users. Our contributions in this paper can be summarized as follows:

- Minimizing simultaneously two contradicting costs, i.e., the channel estimation error and the pilot power consumption, subject to per-user power constraints is a challenging non-convex problem. It can be observed that

Manuscript received January 31, 2020; revised July 08, 2020; accepted August 18, 2020. Date of publication XXX; date of current version XXX. The associate editor coordinating the review of this letter and approving it for publication was Dr. L. Dai. *(Corresponding author: Trinh Van Chien.)*

T. A. Le is with the Department of Design Engineering & Mathematics, Faculty of Science and Technology, Middlesex University, The Burroughs, Hendon, London, NW4 4BT, U. K. Email: t.le@mdx.ac.uk.

T. V. Chien was with the School of Electronics and Telecommunications, Hanoi University of Science and Technology, 100000 Hanoi, Vietnam. He is now with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, L-1855 Luxembourg, Luxembourg. Email: vanchien.trinh@uni.lu.

M. R. Nakhai is with the Department of Engineering, King's College London, London, WC2R 2LS, U. K. Email: reza.nakhai@kcl.ac.uk.

T. Le-Ngoc is with the Department of Electrical and Computer Engineering, McGill University, Montreal, Quebec, H3A 0G4, Canada. Email: tho.le-ngoc@mcgill.ca.

the optimization problem in [1] is a special case of that proposed in this paper. However, the generalization from the problem in [1] to the proposed problem is not trivial. To circumvent the obstacle, we introduce a mixed multi-objective and alternating optimization approach to formulate and solve the problem.

- Although, in general, there is no unique solution to the formulated problem, there exists a set of bounded trade-off *Pareto optimal*[1] solutions. In order to obtain such bounded trade-off Pareto optimal solutions and to characterize the *Pareto frontier*[2] of the proposed multi-objective optimization problem (MOP), a weighted-sum scalarization technique [19] is adopted to transform the proposed MOP into a single objective optimization problem (SOP). The transformed SOP is non-linear and non-convex with respect to the optimization variables, i.e., the desired matrices collecting the pilot signals.
- To deal with both the non-convexity and the non-linearity of the SOP, and to find a distributed solution to the problem, we introduce two auxiliary optimization variables for each cell and adopt the Schur complement [20] to transform the objective function into a linear form while moving the non-linear parts into two new constraints. Since one of the newly introduced constraints is non-convex, we continue with the following two steps. First, we decompose the transformed optimization problem into a number of distributed sub-problems, each can be solved at each cell based on the pilot signal knowledge of the other BSs. Second, we propose an alternating optimization approach to recast each distributed sub-problem into a convex linear matrix inequality (LMI) problem.
- We prove that the proposed alternating optimization approach converges to a Karush-Kuhn-Tucker (KKT) point. Furthermore, we provide a complexity analysis for the proposed LMI optimization problem.
- Numerical results demonstrate that the proposed pilot design outperforms benchmarking schemes thanks to its ability of balancing the total pilot power consumption and the channel estimation error.
- The proposed MOP framework provides a designer (operator) opportunities to select her (his) system's operation points based on her (his) preferences on channel estimation accuracy and power consumption requirements.

The developments in this paper generalize our earlier work in [21] under an MOP framework. More specifically, this generalization leads to further developments outlined above as the first, second and fourth contributions of this paper. These contributions are supported by further analysis and extensive confirmatory simulation results. The optimization problem proposed in this paper differs from that introduced in [1] by four aspects, i.e., *problem formulation aspect* as described in the first contribution; *mathematical aspect* as mentioned in the second, third and forth contributions;[3] *performance aspect* as stated in the fifth contribution; and *application aspect* as pointed out in the sixth contribution. To the best of our knowledge, this is the first work applying MOP for the design of power efficient pilot allocation in cellular massive MIMO systems.

Notation: Bold lower/upper case letters: vectors/matrices; $\|\cdot\|_F$: the Frobenius norm; $\|\cdot\|$: the Euclidean norm; $(\cdot)^H$: complex conjugate transpose operator; $\mathrm{Tr}(\cdot)$: the trace of a matrix; $\mathbf{A} \succeq \mathbf{0}$: $\mathbf{A}$ is a positive semidefinite matrix; $\mathbf{A} \preceq \mathbf{B}$: $\mathbf{B} - \mathbf{A} \succeq \mathbf{0}$; $\mathbf{I}_a$: an $a \times a$ identity matrix; diag{$\mathbf{w}$}: a diagonal matrix with the diagonal entries being elements of the vector $\mathbf{w}$; $(\mathbf{A})^{-1}$: the inverse of a square matrix; $(\mathbf{A})^{\frac{1}{2}}$: the square root of a matrix; $\mathcal{CN}(\cdot;\cdot)$: a circularly symmetric complex Gaussian distribution; $\mathsf{E}[\cdot]$: the expectation of a random variable; $O(\cdot)$: the big-O notation; $\preceq$: element-wise inequality; $(y_i)_{i=1}^{U} : \; y_1 \; y_2 \; \cdots \; y_U \;]^T$; $\mathbf{e}_i$: a suitable size unit column vector containing all zeros except a one at the ith entry; $\mathbf{U}_i$: a suitable size matrix containing all zeros except a one at the ith diagonal entry.

II. System Model

This paper investigates a time-division duplexing multi-cell massive MIMO system with U cells, each comprising of an M-antenna BS and N single-antenna users. The propagation coefficient between the j-th antenna of the BS in cell q, denoted as BS q, and user m in cell p is $\sqrt{\beta_{pmq}}h_{pmqj}$, where β_{pmq} captures the large-scale fading, i.e., path-loss and shadowing, while $h_{pmqj} \sim \mathcal{CN}(0;1)$ represents the small-scale Rayleigh fading. Let $\mathbf{w}_{pm} \in \mathsf{C}^{\tau \times 1}$ denote the pilot signal of τ symbols transmitted by user m in cell p and $\|\mathbf{w}_{pm}\|^2 \leq P_{\max;p}; \forall p$, where $P_{\max;p}$ is the maximum transmit power of each user in cell p allocated to its pilot signal. During the pilot training period, all users synchronously transmit their pilot signals. The received training signal $\mathbf{y}_{qj} \in \mathsf{C}^{\tau \times 1}$ at the j-th antenna element of BS q can be written as:

$$\mathbf{y}_{qj} = \sum_{p=1}^{U} \sum_{m=1}^{N} \sqrt{\beta_{pmq}} h_{pmqj} \mathbf{w}_{pm} + \mathbf{n}_{qj}; \tag{1}$$

where $\mathbf{n}_{qj}$ denotes Gaussian noise with $\mathbf{n}_{qj} \sim \mathcal{CN}(\mathbf{0}; \; ^2\mathbf{I}\;)$. Let

$$\mathbf{Y}_q = [\mathbf{y}_{q1}; \; \mathbf{y}_{q2}; \cdots; \mathbf{y}_{qM}] \in \mathsf{C}^{\tau \times M}; \tag{2}$$

$$\mathbf{N}_q = [\mathbf{n}_{q1}; \; \mathbf{n}_{q2}; \cdots; \; \mathbf{n}_{qM}] \in \mathsf{C}^{\tau \times M}; \tag{3}$$

$$\mathbf{W}_q = [\mathbf{w}_{q1}; \; \mathbf{w}_{q2}; \cdots; \mathbf{w}_{qN}] \in \mathsf{C}^{\tau \times N}; \tag{4}$$

$$_{pq} = \mathrm{diag}\{[\;_{p1q}; \;\; _{p2q}; \cdots; \;\; _{pNq}]^T\} \in \mathsf{C}^{N \times N} \tag{5}$$

compactly represent the received signals, Gaussian noises, pilot signals at all antenna elements of BS q and the corresponding large-scale channel coefficients, respectively. We also stack all small-scale fading channel coefficients of N users in cell p as seen by BS q in the following compact form

$$\mathbf{H}_{pq} = \begin{bmatrix} h_{p1q1}; & \cdots; & h_{p1qM} \\ \vdots & \ddots & \vdots \\ h_{pNq1}; & \cdots; & h_{pNqM} \end{bmatrix} \in \mathsf{C}^{N \times M}: \tag{6}$$

[1] An Pareto optimal is also known as an *e cient optimal*, or a *non-inferior optimal* or a *non-dominated optimal*. A bounded trade-off Pareto optimal, i.e., a properly Pareto optimal, is a Pareto optimal that allows improvements in some objectives with bounded trade-offs from others [19].

[2] The set of objectives corresponding to the set of Pareto optimal solutions is called the Pareto frontier [19].

[3] Only the alternating optimization approach was used in [1].

Then, using (1) – (6), we can express the received training signals by all M antenna elements of BS q as

$$\mathbf{Y}_q = \mathbf{W}_q \sqrt{\rho_{qq}}\mathbf{H}_{qq} + \sum_{p=1;p\neq q} \mathbf{W}_p \sqrt{\rho_{pq}}\mathbf{H}_{pq} + \mathbf{N}_q. \quad (7)$$

The channel estimate $\hat{\mathbf{H}}_{qq}$ of the original channel $\mathbf{H}_{qq}$ can be attained by adopting MMSE estimation [22]:

$$\hat{\mathbf{H}}_{qq} = \mathbb{E}[\mathbf{H}_{qq}\mathbf{Y}_q^H]\ \mathbb{E}[\mathbf{Y}_q\mathbf{Y}_q^H]^{-1}\mathbf{Y}_q. \quad (8)$$

The channel estimation errors at BS q can be represented as $\mathbf{E}_q = \mathbf{H}_{qq} - \hat{\mathbf{H}}_{qq}$. Then the MSE can be expressed as

$$\mathrm{MSE}_q = \mathbb{E}\left[\|\mathbf{E}_q\|_F^2\right] = \mathbb{E}\left[\mathrm{Tr}\ \mathbf{E}_q\mathbf{E}_q^H\right]. \quad (9)$$

Following the similar steps as in [1], one can reformulate (9) as

$$\mathrm{MSE}_q = \varphi_q\ \mathbf{W}_q\ = M\mathrm{Tr}\ \left(\mathbf{I}_N + \sqrt{\rho_{qq}}\mathbf{W}_q^H\mathbf{L}_q^{-1}\mathbf{W}_q\sqrt{\rho_{qq}}\right)^{-1}; \quad (10)$$

where $\mathbf{L}_q = \sum_{p=1;p\neq q}^{U} \mathbf{W}_p \rho_{pq}\mathbf{W}_p^H + \sigma^2\mathbf{I}$.[4]

The total pilot power of all users served by BS q can be expressed as

$$\psi_q\ \mathbf{W}_q\ = \sum_{i=1} \|\mathbf{W}_q\mathbf{e}_i\|^2 = \mathrm{Tr}\ \mathbf{W}_q^H\mathbf{W}_q\ . \quad (11)$$

In the sequel, we adopt an MOP framework to formulate our optimization problem that simultaneously minimizes the channel estimation MSE and the total pilot power.

III. Problem formulation

In this paper, we find the optimal pilot $\mathbf{W}_q$ that simultaneously minimizes the total channel estimation MSE of every BS q, i.e., $\varphi_q\ \mathbf{W}_q$ defined in (10), and the total pilot power of all users of that BS, i.e., $\psi_q\ \mathbf{W}_q$ defined in (11), subject to the power constraint at every user in the network. We first define the objective vector as

$$\mathbf{f}\ \{\mathbf{W}_q\} = \left[\varphi_q\ \mathbf{W}_q\ _{q=1}^{U}\right]^T; \left[\psi_q\ \mathbf{W}_q\ _{q=1}^{U}\right]^T \in \mathbb{R}^{2U}; \quad (12)$$

where $\{\mathbf{W}_q\} = \{\mathbf{W}_1; \mathbf{W}_2; \cdots; \mathbf{W}_U\}$ and the decision space as

$$\mathcal{D} \triangleq \left\{\{\mathbf{W}_q\} \mid \sum_{i=1} \mathbf{U}_i\mathbf{W}_q^H\mathbf{W}_q\mathbf{U}_i \preceq \sum_{i=1} P_{\max;q}\mathbf{U}_i;\ \forall q\right\}; \quad (13)$$

where $P_{\max;q}$ is the maximum pilot power spent by each user in cell q. Hereafter, unless otherwise stated, we assume $i \in \{1; \cdots; N\}$ and $q \in \{1; \cdots; U\}$.

The data of the proposed MOP comprises the decision space $\mathcal{D}$, the objective function vector $\mathbf{f}$, and the objective space $\mathbb{R}^{2U}$. The objective function vectors $\mathbf{f}\ \{\mathbf{W}_q\}$, i.e., $\{\mathbf{W}_q\} \in \mathcal{D}$, are mapped from the objective space $\mathbb{R}^{2U}$ to an ordered space, e.g. $\mathbb{R}^{2U}; \preceq$, where comparisons are performed using the order relation $\preceq$.[5] This mapping is referred to as the model map Φ. The model map Φ describes the link between the objective space and the order space, in which the meaning of the minimization is finally defined [19]. An MOP is completely characterized by data, model map and order space, i.e., $\mathcal{D}; \mathbf{f}; \mathbb{R}^{2U} = = \mathbb{R}^{2U}; \preceq$. Considering an *identity mapping*, i.e., $\Phi(\mathbf{f}) = \mathbf{f}$, we propose the following MOP:

$$\underset{\{\mathbf{W}_q\}\in\mathcal{D}}{\text{minimize}}\quad \mathbf{f}\ \{\mathbf{W}_q\}\ : \quad (14)$$

We continue by recalling the following two definitions [19].

Definition 1: A feasible solution $\{\hat{\mathbf{W}}_q\} \in \mathcal{D}$ is called as *Pareto optimal*, i.e., also known as *e cient*, or *non-inferior* or *non-dominiated*, if and only if there is no other solution $\{\mathbf{W}_q\} \in \mathcal{D}$ such that $\mathbf{f}\ \{\mathbf{W}_q\}\ \preceq \mathbf{f}\ \{\hat{\mathbf{W}}_q\}$.

The Pareto optimal solution $\{\hat{\mathbf{W}}_q\}$ cannot be improved in one of the objectives without adversely degrading at least one other objective. The corresponding objective vector $\mathbf{f}\ \{\hat{\mathbf{W}}_q\}$ is known as the *Pareto dominant vector*.[6] The Pareto optimal set is formed by all Pareto optimal solutions.

Definition 2: The *properly Pareto optimal* solution $\{\tilde{\mathbf{W}}_q\}$ is a Pareto optimal solution with bounded trade-offs amongst the objectives. The corresponding $\mathbf{f}\ \{\tilde{\mathbf{W}}_q\}$ is called *properly Pareto vector*. The collection of all *properly Pareto vectors* is referred to as *Pareto frontier*.

Finding the properly Pareto optimal solutions to attain the Pareto frontier of an MOP class can follow either scalarization or non-scalarization approach. In the former approach, based on the preferential information about objectives, e.g., predefined by the designer (decision) maker, an MOP is then transformed into a SOP. In the latter approach, the priority information about objectives is not available in advance. In such a case, natural inspired or generic algorithms are usually utilized to attain the Pareto frontier by concurrently optimizing all objectives. The scalarization methods obtain the Pareto frontier by iteratively solving several SOPs, where each SOP is generated with given values of priorities amongst these objectives. On the other hand, the non-scalarization methods attain the Pareto frontier by directly solving the MOP. Unfortunately, the non-scalarization methods demand significantly higher computational capacity than their scalarization counterparts.

IV. Pareto-Optimal Pilot Design

We proceed by adopting the weighted sum scalarization method [19] to attain the properly Pareto optimal solutions and the Parto frontier to (14).

Lemma 1: Weighted Sum Method [19]: Let $\omega_q > 0$ and $\tilde{\omega}_q > 0$, $\forall q$ and $\sum_{q=1}^{U} \omega_q + \tilde{\omega}_q = 1$. If $\{\mathbf{W}_q\}^\star$ is the optimal solution to the following SOP

$$\underset{\{\mathbf{W}_q\}\in\mathcal{D}}{\text{minimize}} \sum_{q=1} \omega_q\ \varphi_q\ \{\mathbf{W}_q\} + \sum_{q=1} \tilde{\omega}_q\ \psi_q\ \{\mathbf{W}_q\}\ ; \quad (15)$$

[4]The MSE value in (10) is based on the perfect knowledge of the large-scale-fading coefficients, which is popularly assumed in the Massive MIMO literature, see e.g., [23] and references therein. In [24], the authors demonstrated that the large-scale fading coefficients can be accurately estimated in Massive MIMO communications with a neglect-able performance loss.

[5]A decision space, a.k.a., a variable space, is the space of which the feasible set is a subset whereas an ordered space is the space where comparisons, i.e., element-wise comparisons in this paper, are made between different objective-function vectors.

[6]It is also called as non-inferior vector or non-dominated vector.

then $\{\mathbf{W}_q\}^?$ is also a properly Pareto optimal solution to the MOP class $\mathcal{D}; \mathbf{f}; \mathbb{R}^{2U} \text{ =i.d.= } \mathbb{R}^{2U}; 4$ defined in (14).

It should be noted that $\{\ _1; \ _2; \cdots; \ _U\}$ and $\{\tilde{}_1; \tilde{}_2; \cdots; \tilde{}_U\}$ are predefined values set by the designer (decision maker). By varying those values, the Pareto frontier is attained. Then, the most desirable (suitable) solution to the designer (decision maker) is selected from the Pareto frontier.

Using Lemma 1, we can state the equivalent SOP of (14) as:[7]

$$\begin{aligned} \underset{\{\mathbf{W}_q\}}{\text{minimize}} \quad & \sum_{q=1} \ _q \ _q \mathbf{W}_q + \sum_{q=1} \tilde{}_q \ '_q \ _q \mathbf{W}_q \\ \text{subject to} \quad & \sum_{i=1} \mathbf{U}_i \mathbf{W}_q^H \mathbf{W}_q \mathbf{U}_i \preceq \sum_{i=1} P_{\max;q} \mathbf{U}_i; \ \forall q: \end{aligned} \tag{16}$$

The objective function of problem (16) is non-convex with respect to its optimization variables. To tackle the problem, we first introduce two auxiliary variables $\mathbf{Q}_q$ and $\mathbf{P}_q$ to shift the nonlinear part into constraints and hence transform the objective function into a linear form. Denoting $\{\mathbf{Q}_q\} = \{\mathbf{Q}_1; \cdots; \mathbf{Q}_U\}$, $\{\mathbf{P}_q\} = \{\mathbf{P}_1; \cdots; \mathbf{P}_U\}$, we can rewrite problem (16) in the epigraph form [20, pp.134] as

$$\begin{aligned} \underset{\{\mathbf{W}_q\};\{\mathbf{Q}_q\};\{\mathbf{P}_q\}}{\text{minimize}} \quad & M \sum_{q=1} \ _q \mathrm{Tr}\left(\mathbf{Q}_q + \frac{\tilde{}_q}{_q M}\mathbf{P}_q\right) \\ \text{subject to} \quad & \sum_{i=1} \mathbf{U}_i \mathbf{W}_q^H \mathbf{I}^{-1} \mathbf{W}_q \mathbf{U}_i \preceq \sum_{i=1} P_{\max;q} \mathbf{U}_i; \ \forall q; \\ & \mathbf{I}_N + \ ^{\frac{1}{2}}_{qq} \mathbf{W}_q^H \mathbf{L}_q^{-1} \mathbf{W}_q \ ^{\frac{1}{2}}_{qq} \ ^{-1} \preceq \mathbf{Q}_q; \forall q; \\ & \mathbf{W}_q^H \mathbf{I}^{-1} \mathbf{W}_q \preceq \mathbf{P}_q; \forall q: \end{aligned} \tag{17}$$

Removing the constant M before the sum in the objective function and applying the Schur complement [20], we can equivalently recast problem (17) as

$$\begin{aligned} \underset{\{\mathbf{W}_q\};\{\mathbf{Q}_q\};\{\mathbf{P}_q\}}{\text{minimize}} \quad & \sum_{q=1} \ _q \mathrm{Tr}\left(\mathbf{Q}_q + \frac{\tilde{}_q}{_q M}\mathbf{P}_q\right) \\ \text{subject to} \quad & \sum_{i=1} \begin{bmatrix} P_{\max;q}\mathbf{U}_i & \mathbf{U}_i\mathbf{W}_q^{H^\#} \\ \mathbf{W}_q\mathbf{U}_i & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}; \ \forall q; \\ & \begin{bmatrix} \mathbf{Q}_q & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \ ^{\frac{1}{2}}_{qq} \mathbf{W}_q^H \mathbf{L}_q^{-1} \mathbf{W}_q \ ^{\frac{1}{2}}_{qq} \end{bmatrix} \succeq \mathbf{0}; \forall q; \\ & \begin{bmatrix} \mathbf{P}_q & \mathbf{W}_q^{H^\#} \\ \mathbf{W}_q & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}; \forall q: \end{aligned} \tag{18}$$

Since term $^{\frac{1}{2}}_{qq} \mathbf{W}_q^H \mathbf{L}_q^{-1} \mathbf{W}_q \ ^{\frac{1}{2}}_{qq}$ is nonlinear with respect to the set of optimization variables $\{\mathbf{W}_q\}$, the second set of constraints in (18) is not convex. We take two steps to tackle the problem. Firstly, we introduce a distributed algorithm where each BS q at the t-th iteration optimizes its own pilot signals given the knowledge of the pilot signals of the other BSs in $\mathbf{L}_q^{-1}$ as follows:

$$\begin{aligned} \underset{\mathbf{W}_q^{(t)};\mathbf{Q}_q^{(t)};\mathbf{P}_q^{(t)}}{\text{minimize}} \quad & \ _q \mathrm{Tr}\left(\mathbf{Q}_q^{(t)} + \frac{\tilde{}_q}{_q M}\mathbf{P}_q^{(t)}\right) \\ \text{subject to} \quad & \sum_{i=1} \begin{bmatrix} P_{\max;q}\mathbf{U}_i & \mathbf{U}_i\mathbf{W}_q^{(t);H^\#} \\ \mathbf{W}_q^{(t)}\mathbf{U}_i & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}; \\ & \begin{bmatrix} \mathbf{Q}_q^{(t)} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \ ^{\frac{1}{2}}_{qq} \mathbf{W}_q^{(t);H} \mathbf{L}_q^{-1} \mathbf{W}_q^{(t)} \ ^{\frac{1}{2}}_{qq} \end{bmatrix} \succeq \mathbf{0}; \\ & \begin{bmatrix} \mathbf{P}_q^{(t)} & \mathbf{W}_q^{(t);H^\#} \\ \mathbf{W}_q^{(t)} & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}; \end{aligned} \tag{19}$$

where $\mathbf{W}_q^{(t)}$ is the pilots of cell q attained at the t-th iteration. Furthermore, it is assumed that $\mathbf{L}_q^{-1}$ can be obtained from the previous iteration as

$$(\mathbf{L}_q^{-1})^{(t-1)} = \sum_{p=1;p,q} \mathbf{W}_p^{(t-1)} \ _{pq} \mathbf{W}_p^{(t-1);H} + \ ^2\mathbf{I}; \tag{20}$$

where $\mathbf{W}_p^{(t-1)}$ is the pilots of cell p attained from the $(t-1)$-th iteration. Let $_q = \frac{\tilde{}_q}{_q M}$ and $\mathbf{K}_q^{(t)} = \ ^{\frac{1}{2}}_{qq} \mathbf{W}_q^{(t);H} (\mathbf{L}_q^{-1})^{(t-1)} \mathbf{W}_q^{(t)} \ ^{\frac{1}{2}}_{qq}$. Removing the constant factor $_q$ in the objective function, one can rewrite (19) as

$$\begin{aligned} \underset{\mathbf{W}_q^{(t)};\mathbf{Q}_q^{(t)};\mathbf{P}_q^{(t)}}{\text{minimize}} \quad & \mathrm{Tr}\left(\mathbf{Q}_q^{(t)} + \ _q \mathbf{P}_q^{(t)}\right) \\ \text{subject to} \quad & \sum_{i=1} \begin{bmatrix} P_{\max;q}\mathbf{U}_i & \mathbf{U}_i\mathbf{W}_q^{(t);H^\#} \\ \mathbf{W}_q^{(t)}\mathbf{U}_i & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}; \\ & \begin{bmatrix} \mathbf{Q}_q^{(t)} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{K}_q^{(t)} \end{bmatrix}^\# \succeq \mathbf{0}; \\ & \begin{bmatrix} \mathbf{P}_q^{(t)} & \mathbf{W}_q^{(t);H^\#} \\ \mathbf{W}_q^{(t)} & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}: \end{aligned} \tag{21}$$

Since $\mathbf{L}_q^{-1}$ is known to BS q, the distributed optimization problem (21) now only has $\mathbf{W}_q^{(t)}$, $\mathbf{Q}_q^{(t)}$, and $\mathbf{P}_q^{(t)}$ as optimization variables. However, its second constraint is still not in an LMI form with respect to $\mathbf{W}_q^{(t)}$. We take a second step by introducing Theorem 1 where problem (21) is lower bounded by a LMI program.

Theorem 1: The optimal solution to problem (21) is lower-bounded by the following LMI program

$$\begin{aligned} \underset{\mathbf{W}_q^{(t)};\mathbf{Q}_q^{(t)};\mathbf{P}_q^{(t)};\mathbf{V}_q^{(t)}}{\text{minimize}} \quad & \mathrm{Tr}\left(\mathbf{Q}_q^{(t)} + \ _q \mathbf{P}_q^{(t)}\right) \\ \text{subject to} \quad & \sum_{i=1} \begin{bmatrix} P_{\max;q}\mathbf{U}_i & \mathbf{U}_i\mathbf{W}_q^{(t);H^\#} \\ \mathbf{W}_q^{(t)}\mathbf{U}_i & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}; \\ & \begin{bmatrix} \mathbf{Q}_q^{(t)} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{V}_q^{(t)} \end{bmatrix}^\# \succeq \mathbf{0}; \\ & \begin{bmatrix} \mathbf{P}_q^{(t)} & \mathbf{W}_q^{(t);H^\#} \\ \mathbf{W}_q^{(t)} & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}; \\ & \text{constraint (23)}; \end{aligned} \tag{22}$$

where the newly introduced optimization variable $\mathbf{V}_q^{(t)}$ is

[7] The pilot power constraint for each user is implicitly enforced in (15) by stating $\{\mathbf{W}_q\} \in \mathcal{D}$ while it is explicitly stated as a constraint in (16).

defined as

$$\mathbf{V}_q^{(t)} = \tfrac{1}{2}_{qq}\mathbf{W}_q^{(t);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}_q^{(t-1)}\ \tfrac{1}{2}_{qq} + \tfrac{1}{2}_{qq}\mathbf{W}_q^{(t-1);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}_q^{(t)}\ \tfrac{1}{2}_{qq} - \tfrac{1}{2}_{qq}\mathbf{W}_q^{(t-1);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}_q^{(t-1)}\ \tfrac{1}{2}_{qq}. \quad (23)$$

Proof: Please refer to Appendix B. ∎

Problem (22) can be effectively solved by a standard interior-point method (IPM) [20], [25] provided by optimization packages such as CVX [26]. Algorithm 1 summarizes the procedure to attain the optimal pilot signals for all U cells in the network. In each iteration, the solution $\mathbf{W}_q^{(t)}$ to obtain the MSE for cell l may increase the MSEs of other cells. Consequently, the feasibility checking is necessary to guarantee a non-increasing objective function in problem (18). The convergence of Algorithm 1 is guaranteed by the following Theorem.

Theorem 2: Algorithm 1 generates non-increasing iterations of the objective function in problem (16), characterized as

$$\sum_{q=1}^{U} \quad q\ q\ \mathbf{W}_q^{(t)} + \sum_{q=1}^{U} \tilde{}_q\ '_q\ \mathbf{W}_q^{(t)} \le \sum_{q=1}^{U} \quad q\ q\ \mathbf{W}_q^{(t-1)} + \sum_{q=1}^{U} \tilde{}_q\ '_q\ \mathbf{W}_q^{(t-1)} \quad (24)$$

After a finite number of iterations, Algorithm 1 converges to a KKT point of problem (16).

Proof: Please refer to Appendix C. ∎

It can be observed that the number of decision variables of problem (22) is on the order of $(3N + \ )N$. Let $m = O((3N + \ )N)$. The computational complexity of the major step of Algorithm 1 for any given $\ > 0$ is given by Lemma 2.

Lemma 2: The computational complexities to obtain $\ $-solution to problem (22) is

$$\ln(\ ^{-1})\sqrt{5N+2}\ \ m; \quad (25)$$

where $\ = 11N^3+(6\ +7m)N^2+N\ (6\ +4m)+2\ ^2(m+\ )+m^2$.

Proof: Proof follows the similar steps as in [1], [27]. ∎

Although solving (22) can yield the global optimality per iteration and per cell, it may not be the global optimal solution to the original multicell problem (16) due to its inherent non-convexity. In fact, Algorithm 1 is a suboptimal algorithm with an affordable complexity compared to that of the original non-convex problem in (16).

In (25), $\ln(\ ^{-1})\sqrt{5N+2}\ $ is the iteration complexity to achieve $\ $-solution to problem (22) and $\ m$ is the computation cost per each iteration, see e.g., [25] and [28] for more details. Table I, shown on the top of next page, compares the complexities of the proposed approach and the MSE-Only approach [1]. The information from the table reveals the following facts. The number of decision variables of the proposed approach is N^2 more than that of the MSE-Only approach. Consequently, the computation cost per iteration of the former is higher than that of the later. However, the difference between iteration complexities of the two approaches is marginal indicating the fact that two approaches require almost the same number of iterations to obtain their optimal solutions.

Algorithm 1 Alternating optimization approach for (17)

1: **Inputs:** $\ _{pq}$, $P_{\max;q}$, $\ ^2$, stopping criteria $\ > 0$, initialize $\mathbf{W}_p^{(0)}$, $\forall p; q$; $t = 1$;
2: Each cell q evaluates $\mathbf{L}_q^{(t-1)}$ using (20) and then solves (22) to obtain $\mathbf{W}_q^{(t)}$, $\forall q$; **If** $\mathbf{W}_q^{(t)}$ is feasible to problem (18), **then** broadcast $\mathbf{W}_q^{(t)}$. **Otherwise** keep $\mathbf{W}_q^{(t-1)}$.
3: **If** $\sum_{q=1}^{U}\|\mathbf{W}_q^{(t)} - \mathbf{W}_q^{(t-1)}\|_F \le \ $, **then** Go to step 5;
4: **else if** $\sum_{q=1}^{U}\|\mathbf{W}_q^{(t)} - \mathbf{W}_q^{(t-1)}\|_F > \ $, **then** $t = t + 1$; Go to step 2;
5: **Outputs:** $\mathbf{W}_q^{?} \leftarrow \mathbf{W}_q^{(t)}$, $\forall q$.

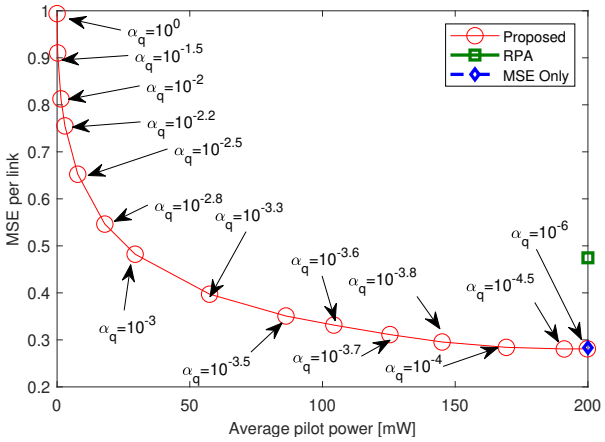

Fig. 1. MSE per BS-user link versus average pilot power.

Overall, the computational complexity of the solution to the proposed problem in (22) is relatively higher than that of our previous work in [1]. However, the extra computation is handled by the BSs rather than the mobile devices that are critically constrained by limited battery power. Furthermore, the extra computational load at BSs comes at the cost for the reduced pilot power, which leads to further power efficiency at mobile devices, as evidenced by our simulation results in Figs. 1 and 4.

Interestingly, observing the parameter $\ $ of two approaches in Table I, one can conclude that as N goes large, the term N^3 will dominate $\ $ and all other terms can be neglected. This is due to the fact that in $\ $ of the proposed approach or the MSE-Only approach, the term with the highest growth rate is $11N^3$ or $10N^3$, respectively. Similarly, N^2 and N terms will dominate in m and the iteration complexity, respectively, as N goes large. Consequently, the complexities of the two approaches become the same as they all converge to the order of $O\ \sqrt{N}N^3O\ N^2\ \approx O\ N^{\frac{11}{2}}$.

V. Simulation Results

A multi-cell Massive MIMO system with wrapped-around is considered for simulations with 4 cells where one BS equipped with 100 antennas is located at the center of each cell to serve 5 users. All users are randomly distributed over the coverage area under a condition that the distance between any user n of

TABLE I
COMPLEXITY COMPARISON

Parameter	Proposed Approach	MSE-Only Approach [1]
Number of decision variables	$(3N + \ )N$	$(2N + \ )N$
Iteration complexity	$\ln(\ ^{-1})\sqrt{5N + 2\ }$	$\ln(\ ^{-1})\sqrt{4N + \ }$
Computation cost per iteration	m	m
	$11N^3 + (6\ \ + 7m)N^2 + N\ (6\ \ + 4m) + 2\ ^2(m + \ ) + m^2$	$10N^3 + (3\ \ + 6m)N^2 + Nm\ (m\ \ + 2) + \ ^2(m + \ ) + m^2$
m	$m = O((3N + \ )N) = O\ 3N^2 + \ N$	$m = O((2N + \ )N) = O\ 2N^2 + \ N$

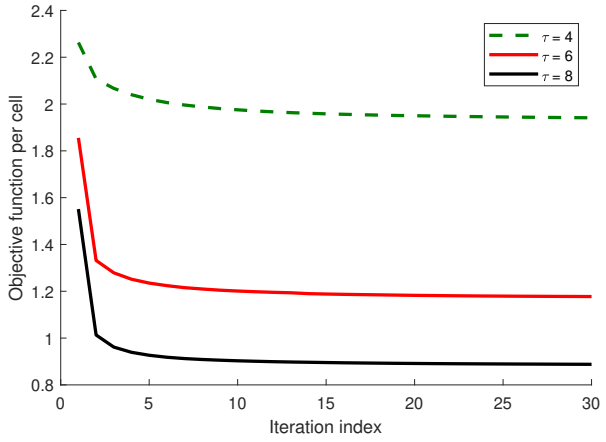

Fig. 2. The convergence of the proposed alternating optimization.

cell p and BS q, denoted as d_{pnq} with $d_{pnq} \geq 0{:}035$ km. The network bandwidth is 20 MHz. The noise variance is -96 dBm for an assumed noise figure of 5 dB. We model the large-scale fading coefficient as

$$_{pnq}[\text{dB}] = -148{:}1 - 37{:}6\log_{10}(d_{pnq}) + z_{pnq}; \qquad (26)$$

where z_{pnq} denotes the shadowing, which follows a log-normal Gaussian distribution with the standard deviation of 7 dB. Monte-Carlo simulations are carried over 200 independent realizations of user locations. Here, we compare the performance of our proposed pilot design with: (*i*) the mean-square-error only (MSE-Only) proposed in [1], and (*ii*) the widely adopted random orthogonal pilot assignment (RPA), e.g., [15], [29], [30]. In the RPA scheme, the orthogonal pilots are shared/reused amongst users in the network and a power of 200 mW is assigned to each pilot symbol. For each user location realization, such pilot signals are obtained by the eigenvectors of a uniformly generated random matrix. The power constraint for pilot signal is set to be $P_{\max;q} = 200$ mW, $\forall q$.

Fig. 1 shows the MSE per BS-user link versus the average pilot power. By varying the weight $_q$, the Pareto frontier of the proposed approach has been obtained and shown as the red curve. Since the MSE-Only and RPA are SOPs, the optimal solutions for them are represented by two points corresponding to the average pilot power of 200 mW in the figure, i.e., the blue-diamond shape for the MSE Only and the green-square shape for the RPA. The Pareto frontier reveals the trade-off between the channel estimation accuracy and the average pilot power via the selection of the weight $_q$. For the weight range from 10^{-6} to 10^{-4}, the proposed approach attains the same MSE as MSE-Only. However, for the weight of 10^{-4}, the proposed approach can save 15:35% on the pilot power consumption compared with its counterpart. It can be observed from the figure that in order to attain the same channel accuracy as the RPA does, the proposed approach operates at the weighted value of 10^{-3} and consumes 85:5 % less power than the RPA. In other words, a reduction of 41:6 % in the channel estimation accuracy, i.e., from MSE of 0:28 to MSE of 0:48, results in the reduction of 85:5 % in the power consumption, i.e., from 200 mW to 29 mW.

Fig. 2 displays the convergence of Algorithm 1 with three different pilot lengths. Hereafter, we set the weighted value at 10^{-5}, i.e., keeping the MSE performance of the proposed approach as same as that of the MSE-Only, when qualifying problem (17) with different pilot lengths. It is clear that the proposed solution method quickly converges with less than 20 iterations. These numerical results verify our mathematical proof in *Theorem 2*. Moreover, a large reduction in the objective function of problem (17) is observed when increasing the pilot length. For example, at the converged point, by adding two more symbols per pilot sequence, i.e., increasing from 4 to 6, the the objective function of problem (17) can be further pushed down 39 % thanks to more degrees of freedom provided.

Fig. 3 illustrates the channel estimation qualities of three approaches, i.e., the proposed, MSE-Only and RPA, versus the length of a pilot sequence . The figure indicates that the proposed approach attains the same MSE level as the MSE-Only counterpart. In the pilot length range from 2 to 4, the MSE performance of the proposed approach is lightly better than that of the RPA. However, when the pilot length increases beyond 4, the proposed approach achieves over 50 % lower MSE than the RPA.

Fig. 4 shows the corresponding pilot power consumption of the proposed and MSE-Only approaches versus the length of a pilot sequence to serve 5 users. It is clear that the proposed approach outperforms its counterpart in terms of lower power consumption. As the pilot length increases from 2 to 5, the power consumption of the proposed and MSE-Only sharply increase from 151:2 mW and 180:5 mW, respectively, to their peaks at 197 mW and 200 mW. When the pilot length increases beyond 5, higher degrees of freedom, offered by more available orthogonal pilot sequences, are available for both approaches to improve their performances. Their improvements, however, appear in two different ways as follows. The MSE-Only approach, having only one objective, searches for solutions to reduce MSE, as seen from Fig. 3, satisfying the power constraints, i.e., utilizing all available power of 200 mW. On

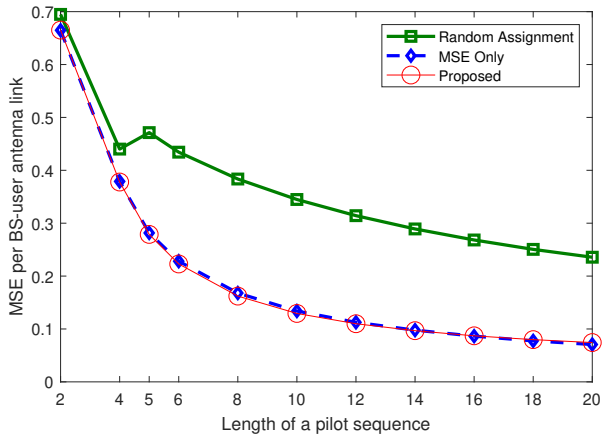

Fig. 3. The MSE per link versus pilot length with 5 users per cell.

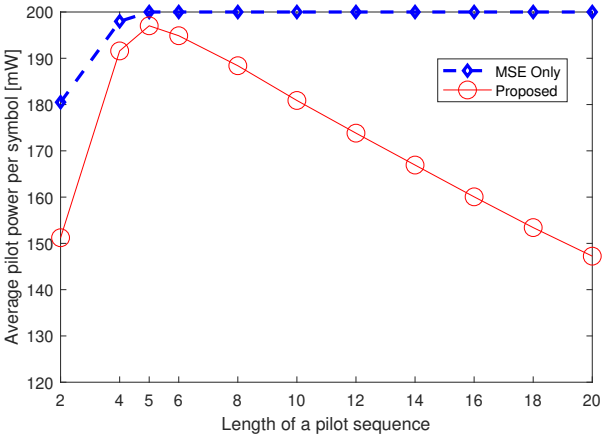

Fig. 4. The average power versus pilot length with 5 users per cell.

the other hand, the proposed approach, having two objectives, can exploit such higher degrees of freedom to further reduce both MSE and its power consumption. The average power of the proposed approach will reach zero (so will the MSEs of the proposed and MSE-Only approach) as the pilot length goes to infinity. This is due to the fact that the feasibility region is significantly enlarged. As a result, the performance gap between the proposed and the MSE-Only increases as the pilot length increases. For instance, jointly optimizing the MSE and pilot power consumption can offer power reduction of only 2:5 % at = 6, but up to 26:4 % at = 20.

It is also worth noticing from Fig. 4 that when the pilot length is comparable with the number of served users, the performance gap between the proposed approach and the MSE-Only is small, i.e., 7 mW, 3 mW, and 5 mW at = 4, = 5, and = 6, respectively. However, compared with the MSE Only, the proposed approach performs extremely well when the pilot length is either smaller or larger than the number of users. For example, the proposed approach consumes 30 mW and 47 mW less power than the MSE-Only at the pilot length of 2 and 18, respectively. This confirms the superior performance of the proposed MOP over a SOP in [1].

Fig. 5 illustrates the performance of the proposed approach

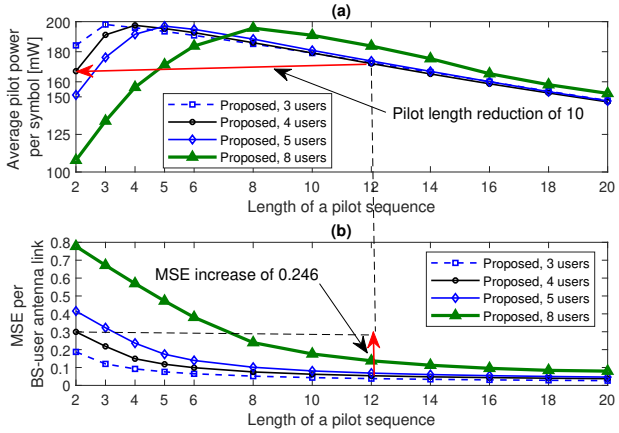

Fig. 5. The average power and MSE per link versus pilot length with different numbers of users per cell.

versus pilot length for different numbers of users per cell. Fig. 5 (a) shows that the power performance curves of the proposed approach for different numbers of users are concave where the maxima appear when the number of users equals to the pilot length. The maximum value of each concave curve is very close to the average pilot power budget per symbol of 200 mW, e.g., 197 mW and 197:5 mW for 5 and 4 users, respectively. Before reaching the maxima for the same pilot length, lower number of users per cell requires more power (Fig. 5 (a)) yet yields lower MSE (Fig. 5 (b)). For instance, at = 2, a 4-users-per-cell scenario consumes 167:1 mW with MSE of 0:3 while a 5-users-per-cell scenario utilizes 151:2 mW with MSE of 0:4157. Further increasing the pilot length beyond the number of users reduces both the average pilot power and MSE. Fig. 5 (a) and (b) also reveal that, for the same pilot power consumption, significantly shorter pilot length can be realized with a trade-off in MSE, e.g., with 4 users per cell, if an MSE increase of 0:246 (from 0:054 to 0:3) is acceptable, then the pilot length can be reduced from 12 to 2 for the same pilot power consumption of 167:1 mW.

VI. Conclusion

We have proposed an MOP approach for pilot design in a multi-cell massive MIMO system. The proposed approach simultaneously minimizes the total channel estimation errors at the BSs and the total pilot power consumed by all mobile users. In order to characterize the Pareto frontier of the proposed MOP, the weighted-sum scalarization method and the alternative optimization technique have been adopted to convert the proposed MOP into distributed SOPs in the form of alternative convex LMI problems to be implemented at each BS in the network. The convergence of the proposed alternative approach has been analytically proved and numerically verified. The Pareto frontier of the proposed MOP, obtained via simulations, provides an insightful view on the trade-off between the channel estimation accuracy and the pilot power consumption for the designers or decision markers to decide the operating points of their systems via the selection of the weighting factors. The Pareto frontier indicates that the

proposed approach outperforms the widely adopted orthogonal pilot design, i.e., the RPA approach, in both lower power consumption and lower channels estimation error. With a proper selection of weighting factor, the proposed approach reduces 15:35% power consumption compared with our recent work in [1] while maintaining the same error performance as its counterpart. Interestingly, the proposed approach still offers significant power reduction, compared with [1], even when the pilot length is less than the number of served users.

Appendix A
Useful Lemmas

Here, we introduce three lemmas that will be used in the proofs of Theorems 1 and 2.

Lemma 3: For a given positive semidefinite matrix $\mathbf{L} \in \mathsf{C}^{\times}$, it holds that

$$\mathbf{X}^H \mathbf{L} \mathbf{X} \succeq \mathbf{0}; \tag{27}$$

for all $\mathbf{X} \in \mathsf{C}^{\times N}$.

Proof: For given $\mathbf{L}$, the proof of (27) is equivalent to prove the following

$$\mathbf{z}^H \mathbf{X}^H \mathbf{L} \mathbf{X} \mathbf{z} \geq 0; \tag{28}$$

for all $\mathbf{z} \in \mathbf{C}^{N\times 1}$. By setting $\mathbf{y} = \mathbf{X}\mathbf{z} \in \mathsf{C}^{\times 1}$, (28) is rewritten as

$$\mathbf{y}^H \mathbf{L} \mathbf{y} \geq 0; \tag{29}$$

which is always true since $\mathbf{L} \succeq \mathbf{0}$. The proof is complete. ∎

Lemma 4: If two positive semidefinite matrices $\mathbf{L}_1$ and $\mathbf{L}_2 \in \mathsf{C}^{N\times N}$ satisfy $\mathbf{L}_1 \succeq \mathbf{L}_2$, then it holds for $\forall \mathbf{Q} \in \mathsf{C}^{N\times K}$ that

$$\mathbf{Q}^H \mathbf{L}_1 \mathbf{Q} \succeq \mathbf{Q}^H \mathbf{L}_2 \mathbf{Q}: \tag{30}$$

Proof: If $\mathbf{L}_1 \succeq \mathbf{L}_2$, then $\mathbf{L}_1 - \mathbf{L}_2 \succeq \mathbf{0}$. Adopting Lemma 3, one can obtain

$$\mathbf{Q}^H (\mathbf{L}_1 - \mathbf{L}_2) \mathbf{Q} \succeq \mathbf{0}: \tag{31}$$

Therefore,

$$\mathbf{Q}^H \mathbf{L}_1 \mathbf{Q} - \mathbf{Q}^H \mathbf{L}_2 \mathbf{Q} \succeq \mathbf{0}: \tag{32}$$

Hence,

$$\mathbf{Q}^H \mathbf{L}_1 \mathbf{Q} \succeq \mathbf{Q}^H \mathbf{L}_2 \mathbf{Q}; \tag{33}$$

which completes the proof. ∎

Lemma 5: For any matrices $\mathbf{A}; \mathbf{B}; \mathbf{C}; \mathbf{D}_1$; and $\mathbf{D}_2 \in \mathsf{C}^{N\times N}$, if $\mathbf{D}_1 \succeq \mathbf{D}_2$, then it holds that

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}_1 \end{bmatrix}^{\#} \succeq \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}_2 \end{bmatrix}^{\#}: \tag{34}$$

Proof: We first assume that (34) is true. Then, for $\forall \mathbf{x} \in \mathsf{C}^{2N\times 1}$, one can state that

$$\mathbf{x}^H \left(\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}_1 \end{bmatrix}^{\#} - \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}_2 \end{bmatrix}^{\#} \right) \mathbf{x} \geq 0: \tag{35}$$

In order to proceed, let us decompose $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2]^T$ where $\mathbf{x}_1; \mathbf{x}_2 \in \mathsf{C}^{N\times 1}$, then plugging $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2]^T$ into (35) with some algebra manipulations, (35) is equivalent to as

$$\mathbf{x}_2^H (\mathbf{D}_1 - \mathbf{D}_2) \mathbf{x}_2 \geq 0; \tag{36}$$

which holds only if $\mathbf{D}_1 \succeq \mathbf{D}_2$. This completes the proof. ∎

Appendix B
Proof of Theorem 1

By observing (20) and adopting Lemma 3, we first conclude that the matrix $(\mathbf{L}_q^{-1})^{(t-1)}$ is positive semidefinite. We then utilize Lemma 3 again with $\mathbf{L} = (\mathbf{L}_q^{-1})^{(t-1)}$ and $\mathbf{X} = \mathbf{W}_q^{(t)} - \mathbf{W}_q^{(t-1)}$ to obtain the property given on the top of next page. Multiplying from left and right of both sides of (39) by $\frac{1}{qq}$ and applying Lemma 4, it holds the fact that

$$\mathbf{K}_q^{(t)} \succeq \mathbf{V}_q^{(t)}: \tag{40}$$

This confirms that the global optimum to problem (21) is an upper bound of (22) as shown in the theorem.

Appendix C
Proof of Theorem 2

The proof consists of the two following main steps. The first step is to approve that an alternating optimization approach in Algorithm 1 produces a non-increasing sequence of the objective function to problem (16). The second step is to manifest that the approximation, described in Theorem 1, produces the KKT point to (16) when Algorithm 1 reaches the convergence.

Observing problem (17), one can conclude that at the optimal point $\{\mathbf{W}_q^?; \mathbf{Q}_q^?; \mathbf{P}_q^?\}$, we have $\mathbf{I}_N + \frac{1}{qq} \mathbf{W}_q^{?;H} \mathbf{L}_q^{-1} \mathbf{W}_q^? \frac{1}{qq}^{-1} = \mathbf{Q}_q^?, \forall q$, $\mathbf{W}_q^{?;H} \mathbf{I}^{-1} \mathbf{W}_q^? = \mathbf{P}_q^?; \forall q$, and $\mathbf{W}_q^?$ is also the optimal solution to problem (16). Hence, according to [20, pp.134], problems (16) and (17) are equivalent. To that end, in the following, we will consider (17) instead of (16).

The first step: We now denote the feasible set of (17) as $\mathcal{F}$ that contains all possibilities of $\{\mathbf{W}_q; \mathbf{Q}_q; \mathbf{P}_q\}$ satisfying the constraints of this problem. We also denote $\mathcal{F}_q^{(t)}$ is the feasible set of problem (22) at the t-th iteration, where $\mathcal{F}_q^{(t)} \subset \mathcal{F}$. Let $\mathcal{I}_q^{(t)}$ contain the optimal solution to problem (22). By applying Lemma 5, the lower bound on the second constraint of (22) is obtained as

$$\begin{bmatrix} \mathbf{Q}_q^{(t)} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{K}_q^{(t)} \end{bmatrix}^{\#} \succeq \begin{bmatrix} \mathbf{Q}_q^{(t)} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{V}_q^{(t)} \end{bmatrix}^{\#}: \tag{41}$$

We stress that (41) holds for every cell. This property implies that the global solution to (22) is always feasible to (21). We now denote $f_q^{(t)}$ the objective function of problem (22). The bound in (41) produces the following sequence of inequalities as

$$f_q^{(t)} \left(\mathcal{I}_q^{(t-1)} \right) \overset{(a)}{\geq} f_q^{(t)} \left(\mathcal{F}_q^{(t)} \right) \overset{(b)}{\geq} f_q^{(t)} \left(\mathcal{I}_q^{(t+1)} \right); \tag{42}$$

where $\mathcal{I}_q^{(t-1)}$ and $\mathcal{I}_q^{(t-1)}$ contain the optimal solutions to problem (22) at the $(t-1)$−th iteration and the $(t+1)$−th iteration, respectively. In (42), (a) is obtained by the fact that $\mathcal{I}_q^{(t-1)} \subseteq \mathcal{F}_q^{(t)}$ and therefore the optimal solution to problem (22) is only a feasible point to problem (21). Meanwhile, (b) is obtained since solving problem (22) always finds the best solution in the feasible set $\mathcal{F}_q^{(t)}$. We notice that (42) creates a non-increasing function along iterations while minimizing problem (22) to seek for a solution to problem (21). Repeating this sequence of inequalities to all U cell and thanks to the

$$\mathbf{W}^{(t)} - \mathbf{W}^{(t-1)\ H}(\mathbf{L}_q^{-1})^{(t-1)}\ \mathbf{W}^{(t)} - \mathbf{W}^{(t-1)}\ \succeq \mathbf{0} \tag{37}$$

$$\Leftrightarrow \mathbf{W}^{(t);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}^{(t)} - \mathbf{W}^{(t);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}^{(t-1)} - \mathbf{W}^{(t-1);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}^{(t)} + \mathbf{W}^{(t-1);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}^{(t-1)} \succeq \mathbf{0} \tag{38}$$

$$\Leftrightarrow \mathbf{W}^{(t);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}^{(t)} \succeq \mathbf{W}^{(t);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}^{(t-1)} + \mathbf{W}^{(t-1);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}^{(t)} - \mathbf{W}^{(t-1);H}(\mathbf{L}_q^{-1})^{(t-1)}\mathbf{W}^{(t-1)}: \tag{39}$$

back tracking condition in Algorithm 1, we can approve the non-increasing property in (24).

The second step: Algorithm 1 must converge to a fixed point since the feasible domain to each optimization problem is a convex set hence compact.

Let f be the objective function of problem (17). At the fixed point which is obtained at the t^*-iteration, we have the following properties

$$\frac{@f\ \mathcal{I}_q^{(t^*)}}{@\mathbf{W}_q^{(t^*)}} = \frac{@f_q^{(t^*)}\ \mathcal{I}_q^{(t^*)}}{@\mathbf{W}_q^{(t^*)}}; \tag{43}$$

$$\frac{@f\ \mathcal{I}_q^{(t^*)}}{@\mathbf{Q}_q^{(t^*)}} = \frac{@f_q^{(t^*)}\ \mathcal{I}_q^{(t^*)}}{@\mathbf{Q}_q^{(t^*)}}; \tag{44}$$

$$\frac{@f\ \mathcal{I}_q^{(t^*)}}{@\mathbf{P}_q^{(t^*)}} = \frac{@f_q^{(t^*)}\ \mathcal{I}_q^{(t^*)}}{@\mathbf{P}_q^{(t^*)}}; \tag{45}$$

which leads to the following inequalities

$$\mathrm{Tr}\left[\left[\frac{@f\ \mathcal{I}_q^{(t^*)}}{@\mathbf{W}_q^{(t^*)}}\right]^H\ \mathbf{W}_q - \mathbf{W}_q^{(t^*)}\ \right] \geq 0; \tag{46}$$

$$\mathrm{Tr}\left[\left[\frac{@f\ \mathcal{I}_q^{(t^*)}}{@\mathbf{Q}_q^{(t^*)}}\right]^H\ \mathbf{Q}_q - \mathbf{Q}_q^{(t^*)}\ \right] \geq 0; \tag{47}$$

$$\mathrm{Tr}\left[\left[\frac{@f\ \mathcal{I}_q^{(t^*)}}{@\mathbf{P}_q^{(t^*)}}\right]^H\ \mathbf{P}_q - \mathbf{P}_q^{(t^*)}\ \right] \geq 0; \tag{48}$$

so that the obtained solution is a stationary point. Furthermore, the KKT point of each optimization problem (22) contributes to the KKT point of problem (17) when we use (43)–(45) to do a matching process similarly to what was done in Theorem 1 of [31]. The proof is completed.

References

[1] H. Al-Salihi, T. V. Chien, T. A. Le, and M. R. Nakhai, "A successive optimization approach to pilot design for multi-cell Massive MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 1086–1089, 2018.

[2] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO-5G system," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 905–923, 2016.

[3] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[4] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 942–953, Oct. 2014.

[5] L. You, X. Gao, A. L. Swindlehurst, and W. Zhong, "Channel acquisition for massive MIMO-OFDM with adjustable phase shift pilots," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1461–1476, Mar. 2016.

[6] X. Luo, X. Zhang, P. Cai, and H. Qian, "Aligning power in multiple domains for pilot decontamination in massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 7919–7935, Dec. 2017.

[7] S. Ma, E. L. Xu, A. Salimi, and S. Cui, "A novel pilot assignment scheme in massive MIMO networks," *IEEE Wireless Communications Letters*, vol. 7, no. 2, pp. 262–265, Apr. 2018.

[8] Y. Gu and Y. D. Zhang, "Pilot design for gaussian mixture channel estimation in massive MIMO," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 3266–3270.

[9] W. Zhang and W. Zhang, "On optimal training in massive MIMO systems with insufficient pilots," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[10] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 787–801, Oct. 2014.

[11] T. C. Mai, H. Q. Ngo, M. Egan, and T. Q. Duong, "Pilot power control for cell-free massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 11 264–11 268, Nov. 2018.

[12] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[13] S. Bazzi and W. Xu, "Downlink training sequence design for FDD multiuser massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 18, pp. 4732–4744, Sep. 2017.

[14] N. Akbar, N. Yang, P. Sadeghi, and R. A. Kennedy, "Multi-cell multiuser massive MIMO networks: User capacity analysis and pilot design," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5064–5077, Dec. 2016.

[15] T. V. Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell Massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000 – 2015, 2018.

[16] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P. M. Grant, H. Haas, J. S. Thompson, I. Ku, C. Wang, T. A. Le, M. R. Nakhai, J. Zhang, and L. Hanzo, "Green radio: radio techniques to enable energy-efficient wireless networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 46–54, May 2011.

[17] A. Abrol and R. K. Jha, "Power optimization in 5G networks: A step towards GrEEn communication," *IEEE Access*, vol. 4, pp. 1355–1374, Apr. 2016.

[18] G. Dong, H. Zhang, S. Jin, and D. Yuan, "Energy-efficiency-oriented joint user association and power allocation in distributed massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5794–5808, Jun. 2019.

[19] M. Ehrgott, *Multicriteria Optimization*. Springer, 2005.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[21] T. A. Le, T. Van Chien, and M. R. Nakhai, "A power efficient pilot design for multi-cell massive MIMO systems," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov. 2018, pp. 823–827.

[22] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[23] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[24] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell massive mimo systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2363–2378, 2016.

[25] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and C.-Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO