Enhanced Channel Estimation in Massive MIMO via Coordinated Pilot Design

Kaiming Shen, *Member, IEEE*, Hei Victor Cheng, *Member, IEEE*,
Xihan Chen, *Student Member, IEEE*, Yonina C. Eldar, *Fellow, IEEE*, and Wei Yu, *Fellow, IEEE*

Abstract—Pilot contamination is a limiting factor in multicell massive multiple-input multiple-output (MIMO) systems because it can severely impair channel estimation. Prior works have suggested coordinating pilot design across cells in order to reduce the channel estimation error caused by pilot contamination. In this paper, we propose a method for coordinated pilot design using fractional programming to minimize the weighted mean squared-error (MSE) in channel estimation. In particular, we apply the recently proposed quadratic transform to the MSE expression which allows the effect of pilot contamination to be decoupled. The resulting problem reformulation enables the pilots to be optimized in closed form if they can be designed arbitrarily. When the pilots are restricted to a given set of orthogonal sequences, pilot optimization reduces to an assignment problem which can be solved by weighted bipartite matching. Furthermore, we consider the max-min fairness of data rates with orthogonal pilots and obtain an extension of the proposed method to correlated Rayleigh fading. Finally, simulations demonstrate the advantage of the proposed (orthogonal and nonorthogonal) pilot designs as compared with state-of-the-art methods in combating pilot contamination.

Index Terms—Pilot contamination, massive MIMO systems, weighted MSE minimization, max-min fairness of rates, orthogonal and nonorthogonal pilot designs, correlated Rayleigh fading.

I. INTRODUCTION

ACQUISITION of channel state information (CSI) is crucial in massive multiple-input multiple-output (MIMO)

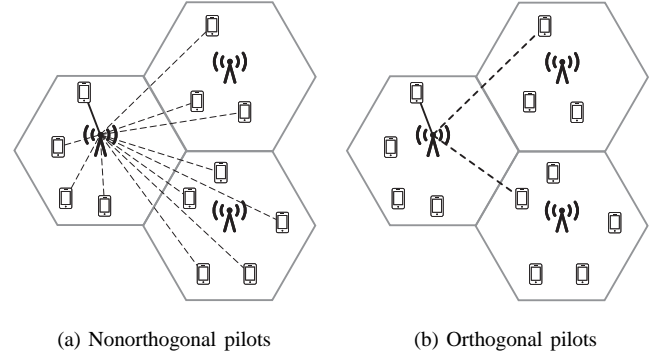
(a) Nonorthogonal pilots   (b) Orthogonal pilots

Fig. 1. Orthogonal pilots versus nonorthogonal pilots. The solid line is the desired pilot while the dashed lines are the interfering pilots; the width of the dashed lines is proportional to the correlation with the desired pilot.

wireless networks. A main challenge in channel estimation is that due to the limited coherence time, pilot sequences assigned to multiple users across multiple cells cannot all be orthogonal. The nonorthogonality between the pilots, e.g., when the same set of pilots is reused across cells, causes the channel estimation for each user to be affected by the pilots of other users. This is referred to in the literature as pilot contamination [2], [3].

This work pursues a strategy of designing pilot sequences of user terminals across cells as a function of their large-scale fading (assuming that user terminals are relatively stationary) in order to minimize pilot contamination. Following the recent works of [4], [5], the idea is that the effect of pilot contamination mainly depends on the large-scale fading between user terminals and base stations (BSs). For example, if some interfering pilot signal is weak, then the desired pilots can afford to have higher correlation with it. Thus, judicious pilot design for the different users across multiple cells can help alleviate the pilot contamination effect.

The above goal can be further characterized as minimizing some suitable system-level metric of channel estimation performance by choosing the pilot sequences properly. The authors in [4], [5] consider the minimum mean squared-error (MMSE) as the error metric. Here we additionally include weights, each reflecting the extent to which a particular user is affected by pilot contamination; thus weaker users may be assigned higher weights. We begin with the nonorthogonal case as illustrated in Fig. 1(a). Pilot design in this setting entails solving a multidimensional nonconvex problem. In contrast to standard tools such as greedy methods [4], [5]

K. Shen was with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada. He is now with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China (e-mail: shenkaiming@cuhk.edu.cn)

H. V. Cheng and W. Yu are with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: hei.cheng@utoronto.ca, weiyu@ece.utoronto.ca).

X. Chen is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: chenxihan@zju.edu.cn).

Y. C. Eldar is with the Faculty of Mathematics and Computer Science, Weizmann institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

and successive optimization [6], our approach is tailored to the fractional structure of the nonorthogonal pilot design problem. Specifically, minimizing the weighted MMSE using arbitrary pilots can be interpreted as a continuous sum-of-ratios programming problem. We simplify the problem by separating the numerator and denominator of each ratio. We achieve this separation by using the *quadratic transform* [7], [8], which is capable of decoupling more than one ratio. Earlier approaches to fractional programming (FP) such as the Dinkelbach's method [9], [10] cannot perform such a separation.

Although nonorthogonal pilots provide more accurate channel estimation, an orthogonal pilot scheme shown in Fig. 1(b) may still be favored in practice owing to its simple implementation. The assignment of orthogonal pilots to users involves a challenging combinatorial optimization. In comparison to the state-of-the-art method [11] that assigns the orthogonal pilots to one cell at a time, we show that by using our decoupling approach, coordinated pilot design may be reformulated as a multi-cell assignment problem which can be efficiently solved via weighted bipartite matching.

While [12]–[14] consider blind channel estimation without using pilots, it is common in the literature of massive MIMO to enhance channel estimation accuracy via pilot optimization. The allocation of fixed orthogonal pilots across users is a well-studied problem in this area. In order to mitigate the pilot contamination effect, [15] proposes assigning pilots based on the orthogonality between the users' channels, which is quantified by the angle of arrival (AoA) of the received signal. Alternatively, [11] suggests a greedy method that optimizes the pilot assignment for one cell at a time. Another common heuristic [16], [17] is to partially reuse pilots among cell-edge users; pilots are fully reused only among those center users that are more resistant to pilot contamination. In contrast to the above works assuming orthogonal pilots with fixed powers, a more sophisticated pilot design [18]–[22] incorporates power control into the orthogonal pilot assignment. As a further extension, [23] allows each user to combine multiple orthogonal pilots, but the resulting pilot sequences cannot be arbitrary as discussed in Section II-A. Arbitrary nonorthogonal pilots are considered in [4], [5], [24]. A multi-cell precoding scheme is used in [24] to combat pilot contamination, while [4], [5] design the pilot symbols sequentially in a greedy fashion. In contrast to these existing works that consider only a particular type of pilot design, the FP-based framework proposed here can be used for both the orthogonal and nonorthogonal cases.

Prior works on pilot design can also be categorized according to their objective functions. The MSE-based metric has been considered extensively in the literature, including the sum-of-MSEs minimization [4], [5], [15] and the sum-of-normalized-MSEs minimization [20]–[22], [25]. This work considers a general sum-of-weighted-MSEs minimization; the extension for correlated Rayleigh fading is also studied. Achievable rate is another common metric, e.g., [18] considers maximizing an increasing concave utility function of rates, [19] aims to minimize the total power consumption under the rate constraint, and [11], [23] seek max-min fairness across user rates. In Section V-D we show that max-min-rate

optimization can be addressed from an FP perspective as well. We further provide a closed-form rate expression for arbitrary nonorthogonal pilots, whereas the result in [23] is for pilots which are nonnegative combinations of some fixed orthogonal sequences.

The main contributions of this work include:

- *Unified framework for pilot design:* The existing works mostly consider a particular type of pilot design. Based on our proposed approach of viewing weighted MSE minimization from an FP perspective, this work suggests a unified framework that accounts for both orthogonal pilot design and nonorthogonal pilot design.
- *Achievable rate analysis:* The achievable rate of massive MIMO is typically considered for orthogonal pilots; the recent work [23] gives an extension for a special type of nonorthogonal pilots. This paper further generalizes the closed-form rate expression to arbitrary pilots.
- *Max-min fairness of user rates:* The well-known "smart pilot assignment" in [11] aims to maximize achievable rates with max-min fairness. We in addition consider the optimal pilot powers assuming that a given set of normalized orthogonal pilots have been assigned. It turns out that the power control problem, though nonconvex, can be efficiently solved via max-min-ratio FP.
- *Correlated channel case:* The MSE term has a matrix ratio form in the presence of correlated Rayleigh fading. We show that the numerator and denominator of each matrix ratio can still be decoupled by using a recent technique in [26]. As a result, the proposed FP framework for pilot design can be readily extended to the correlated channel case.

The rest of the paper is organized as follows. Section II describes the massive MIMO system and formulates the pilot design problem. Section III briefly reviews the quadratic transform—a new FP technique [7], [26]. Section IV examines the nonorthogonal pilot design while Section V treats the orthogonal setting. Section VI extends the results to correlated channel estimation. Section VII analyzes both computational complexity and communication complexity for the proposed algorithms. Numerical results are presented in Section VIII. Finally, Section IX concludes the paper.

Throughout the paper we use the following notation. We use $\|\cdot\|$ to denote the Euclidean norm, $(\cdot)^\top$ the transpose, $(\cdot)^H$ the conjugate transpose, $\mathrm{vec}(\cdot)$ the vectorization, $\mathrm{tr}(\cdot)$ the trace. We let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}_+$ the set of nonnegative numbers, $\mathbb{C}^{m\times n}$ the $m\times n$ dimensional complex space, $\mathbb{H}^{m\times m}$ the set of $m\times m$ Hermitian matrices. In addition, $\Re$ is the real part of a complex number, $\boldsymbol{I}_n$ is an $n\times n$ identity matrix, $[1:n]$ is the discrete set $\{1,2,\ldots,n\}$, $\mathbf{e}_m^n$ is an $m\times 1$ all-zeros vector except its n entry being 1, and $\mathbf{E}_m^{[n_1:n_2]}$ is an $m\times(n_2-n_1+1)$ matrix $[\mathbf{e}_m^{n_1},\mathbf{e}_m^{n_1+1},\ldots,\mathbf{e}_m^{n_2}]$. We use underline to denote a collection of variables, e.g., $\underline{\mathbf{X}}=\{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n\}$. For ease of reference, we list the main variables in Table I.

TABLE I
LIST OF MAIN VARIABLES

Notation	Definition
M	number of antennas at each BS
L	number of cells
K	number of user terminals per cell
τ	length of pilot
l, i	index of BS or cell
k, j	index of user terminal in the cell
$\boldsymbol{\phi}_{lk}$	orthogonal or nonorthogonal pilot of user (l, k)
$\boldsymbol{\varphi}_s$	sth possible normalized orthogonal pilot
$\boldsymbol{\psi}_{lk}$	normalized orthogonal pilot assigned to user (l, k)
p_{lk}	transmit power for $\boldsymbol{\psi}_{lk}$
β_{ilk}	large-scale fading from user (l, k) to BS i
$\mathbf{R}_{ilk}$	covariance matrix of from user (l, k) to BS i

II. SYSTEM MODEL

A. Pilot Design Settings

Consider an uplink massive MIMO system with L cells, each cell consisting of one BS and K user terminals. Assume that every BS has M antennas and every user terminal has a single antenna. The full coherence bandwidth is reused across the cells. We use (l, k) to index the kth user in the lth cell, for $l \in [1 : L]$ and $k \in [1 : K]$; another index (i, j) is similarly defined. Let $\mathbf{h}_{lij} \in \mathbb{C}^M$ be the uplink channel from user (i, j) to BS l. Each channel is modeled as

$$\mathbf{h}_{lij} = \sqrt{\beta_{lij}} \mathbf{g}_{lij}, \tag{1}$$

in which the large-scale fading β_{lij} is known *a priori* while the Rayleigh fading $\mathbf{g}_{lij}$ is drawn i.i.d. from a complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \boldsymbol{I}_M)$. We begin with the above uncorrelated channel model. An extension with correlated $\mathbf{g}_{lij}$ is provided in Section VI.

Every pilot sequence consists of τ symbols. Assume that each channel $\mathbf{h}_{lij}$ is fixed throughout the pilot sequence. Let $\boldsymbol{\phi}_{lk} \in \mathbb{C}^\tau$ be the pilot sequence of user (l, k). The received pilot signal $\mathbf{Y}_l \in \mathbb{C}^{M \times \tau}$ at BS l can be expressed as

$$\mathbf{Y}_l = \sum_{(i,j)} \mathbf{h}_{lij} \boldsymbol{\phi}_{ij}^\top + \mathbf{Z}_l, \tag{2}$$

where the additive background noise $\mathbf{Z}_l \in \mathbb{C}^{M \times \tau}$ has each entry drawn i.i.d. from $\mathcal{CN}(0, \sigma^2)$. We compare three types of pilot design as follows:

1) Orthogonal Pilots: Each pilot $\boldsymbol{\phi}_{lk}$ is structured as

$$\boldsymbol{\phi}_{lk} = \sqrt{p_{lk}} \boldsymbol{\psi}_{lk} \quad \text{with} \quad 0 < p_{lk} \leq P_{\max}, \tag{3}$$

where $P_{\max}$ is the power constraint, and $\boldsymbol{\psi}_{lk}$ is selected from a given set of normalized orthogonal pilots $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_\tau\}$ with each $\|\boldsymbol{\varphi}_s\|^2 = \tau$. In particular, the convention requires that the users in the same cell be assigned different pilots, e.g., $\boldsymbol{\psi}_{lk} \neq \boldsymbol{\psi}_{lk'}$ for $k \neq k'$. This type of pilot design is common in the existing literature.

2) Restricted Nonorthogonal Pilots: The recent work [23] proposes that each user (l, k) sends a nonnegative combination of some fixed orthogonal sequences:

$$\boldsymbol{\phi}_{lk} = \sum_{s=1}^\tau c_{lk}^{(s)} \boldsymbol{\varphi}_s \quad \text{with} \quad \sum_{s=1}^\tau \left| c_{lk}^{(s)} \right|^2 \leq P_{\max} \text{ and } c_{lk}^{(s)} \in \mathbb{R}_+. \tag{4}$$

Since each $c_{lk}^{(s)}$ is nonnegative, the choice of pilot is restricted to the positive orthant of the linear space spanned by the normalized orthogonal pilots $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_\tau\}$. This assumption is critical to the geometric programming method in [23].

3) Nonorthogonal Pilots: A further generalization allows each $\boldsymbol{\phi}_{lk}$ to be an arbitrary sequence in the τ-dimensional space under the power constraint:

$$\boldsymbol{\phi}_{lk} \in \mathbb{C}^\tau \quad \text{with} \quad \|\boldsymbol{\phi}_{lk}\|^2 \leq \tau P_{\max}. \tag{5}$$

This general form of pilots has been studied in [4], [5], [24]. Note that (4) is equivalent to (5) when each $c_{lk}^{(s)}$ can be an arbitrary complex number not restricted to $\mathbb{R}_+$.

B. Weighted MSE Minimization

Based on the received pilot signal $\mathbf{Y}_l$, each BS l aims to recover its own channels $\{\mathbf{h}_{ll1}, \ldots, \mathbf{h}_{llK}\}$. The channel estimate of $\mathbf{h}_{llk}$ is chosen to minimize the MSE, i.e.,

$$\hat{\mathbf{h}}_{llk} = \arg\min_{\mathbf{h}} \mathbb{E}\left[\|\mathbf{h}_{llk} - \mathbf{h}\|^2\right], \tag{6}$$

where the expectation is taken over Rayleigh fading $\mathbf{g}$. Following the standard steps as shown in [4], [5], we obtain the MMSE estimator at BS l:

$$\hat{\mathbf{h}}_{llk} = \left(\beta_{llk} \boldsymbol{\phi}_{lk}^H \otimes \boldsymbol{I}_M\right)\left(\mathbf{D}_l \otimes \boldsymbol{I}_M\right)^{-1} \text{vec}(\mathbf{Y}_l), \tag{7}$$

where $\mathbf{D}_l$ is the covariance matrix of $\mathbf{Y}_l$, i.e.,

$$\mathbf{D}_l = \sigma^2 \boldsymbol{I}_\tau + \sum_{(i,j)} \beta_{lij} \boldsymbol{\phi}_{ij} \boldsymbol{\phi}_{ij}^H. \tag{8}$$

The corresponding MSE of user (l, k) is

$$\text{MSE}_{lk} = M\beta_{llk} - M\beta_{llk}^2 \left(\boldsymbol{\phi}_{lk}^H \mathbf{D}_l^{-1} \boldsymbol{\phi}_{lk}\right). \tag{9}$$

In contrast, the MMSE estimator of $\mathbf{h}_{llk}$ in [23] is suboptimal because it is based on a partial projection $\mathbf{Y}_l \boldsymbol{\phi}_{lk}$ rather than the complete received signal $\mathbf{Y}_l$; it attains the minimum MSE in (9) only when the orthogonal pilot design is assumed.

Given a set of positive weights $\alpha_{lk} > 0$, we seek a set of pilots that lead to the minimum weighted sum MSE of channel estimation throughout the multicell system, i.e.,

$$\underset{\underline{\boldsymbol{\phi}}}{\text{minimize}} \quad \sum_{(l,k)} \alpha_{lk} \text{MSE}_{lk}. \tag{10}$$

The MSE weights α_{lk} are chosen on a case-by-case basis. For instance, we may set $\alpha_{lk} = 1$ to minimize the sum of MSEs [5], or $\alpha_{lk} = 1/\beta_{llk}$ to minimize the sum of normalized MSEs [20]–[22], [25].

With (9) substituted in (10) and some constant terms removed, the above problem can be converted to

$$\underset{\underline{\boldsymbol{\phi}}}{\text{maximize}} \quad \sum_{(l,k)} \alpha_{lk} \beta_{llk}^2 \left(\boldsymbol{\phi}_{lk}^H \mathbf{D}_l^{-1} \boldsymbol{\phi}_{lk}\right) \tag{11a}$$

$$\text{subject to} \quad \|\boldsymbol{\phi}_{lk}\|^2 \leq P_{\max}. \tag{11b}$$

Here we assume that the pilots can be arbitrarily designed. If an orthogonal pilot scheme is used, then an additional constraint (3) is included in (11). As a result, the problem involves the assignment of (normalized) orthogonal pilots $\{\varphi_1, \ldots, \varphi_\tau\}$ along with power control of p_{lk}.

III. QUADRATIC TRANSFORM

Our proposed approach to the pilot design problem in (11) is based on the quadratic transform [7], [26]. This technique is capable of decoupling multiple (matrix) ratios simultaneously, whereas the traditional Dinkelbach's method [9], [10] is restricted to a single scalar ratio.

Theorem 1 (Quadratic Transform [7]): Given a nonempty constraint set $\mathcal{X}$ as well as N tuples of function $\mathbf{a}_n(\mathbf{x}) \in \mathbb{C}^m$, function $\mathbf{B}_n(\mathbf{x}) \in \mathbb{H}^{m \times m}$, and nondecreasing function $f_n : \mathbb{R}_+ \mapsto \mathbb{R}$, for $n \in [1 : N]$, the sum-of-functions-of-ratio problem

$$\underset{\mathbf{x}}{\text{maximize}} \quad \sum_{n=1}^{N} f_n\Big(\mathbf{a}_n^H(\mathbf{x})\mathbf{B}_n^{-1}(\mathbf{x})\mathbf{a}_n(\mathbf{x})\Big) \qquad (12a)$$

$$\text{subject to} \quad \mathbf{x} \in \mathcal{X} \qquad (12b)$$

is equivalent to

$$\underset{\mathbf{x}, \underline{\boldsymbol{\lambda}}}{\text{maximize}} \quad \sum_{n=1}^{N} f_n\Big(2\Re\{\mathbf{a}_n^H(\mathbf{x})\boldsymbol{\lambda}_n\} - \boldsymbol{\lambda}_n^H \mathbf{B}_n(\mathbf{x})\boldsymbol{\lambda}_n\Big) \quad (13a)$$

$$\text{subject to} \quad \mathbf{x} \in \mathcal{X} \qquad (13b)$$

$$\boldsymbol{\lambda}_n \in \mathbb{C}^m, \qquad (13c)$$

where $\boldsymbol{\lambda}_n$ is an auxiliary variable introduced for each ratio term $\mathbf{a}_n^H(\mathbf{x})\mathbf{B}_n^{-1}(\mathbf{x})\mathbf{a}_n(\mathbf{x})$.

The equivalence between (12) and (13) no longer holds when the argument of $f_n(\cdot)$ is a negative ratio. As shown in [26], the quadratic transform amounts to constructing a *surrogate function* so that the original objective function is lower bounded by the new objective function given any $\underline{\boldsymbol{\lambda}}$; this can be guaranteed if and only if the value of each $\mathbf{a}_n^H(\mathbf{x})\mathbf{B}_n^{-1}(\mathbf{x})\mathbf{a}_n(\mathbf{x})$ is positive.

The quadratic transform can be further extended to the matrix ratio case as stated in the following theorem.

Theorem 2 (Matrix Quadratic Transform [26]): Given a nonempty constraint set $\mathcal{X}$ as well as N tuples of functions $\mathbf{A}_n(\mathbf{x}) \in \mathbb{C}^{m_1 \times m_2}$, functions $\mathbf{B}_n(\mathbf{x}) \in \mathbb{H}^{m_1 \times m_1}$, and nondecreasing functions $F_n : \mathbb{H}^{m_2 \times m_2} \mapsto \mathbb{R}$ in the sense that $F_n(\mathbf{C}) \geq F_n(\mathbf{C}')$ if $\mathbf{C} \succeq \mathbf{C}'$, for $n \in [1 : N]$, the sum-of-functions-of-matrix-ratio problem

$$\underset{\mathbf{x}}{\text{maximize}} \quad \sum_{n=1}^{N} F_n\Big(\mathbf{A}_n^H(\mathbf{x})\mathbf{B}_n^{-1}(\mathbf{x})\mathbf{A}_n(\mathbf{x})\Big) \qquad (14a)$$

$$\text{subject to} \quad \mathbf{x} \in \mathcal{X} \qquad (14b)$$

is equivalent to

$$\underset{\mathbf{x}, \underline{\boldsymbol{\Lambda}}}{\text{maximize}} \quad \sum_{n=1}^{N} F_n\Big(2\Re\{\mathbf{A}_n^H(\mathbf{x})\boldsymbol{\Lambda}_n\} - \boldsymbol{\Lambda}_n^H \mathbf{B}_n(\mathbf{x})\boldsymbol{\Lambda}_n\Big)$$
$$\qquad (15a)$$

$$\text{subject to} \quad \mathbf{x} \in \mathcal{X} \qquad (15b)$$

$$\boldsymbol{\Lambda}_n \in \mathbb{C}^{m_1 \times m_2}, \qquad (15c)$$

where $\boldsymbol{\Lambda}_n$ is an auxiliary variable introduced for each matrix ratio term $\mathbf{A}_n^H(\mathbf{x})\mathbf{B}_n^{-1}(\mathbf{x})\mathbf{A}_n(\mathbf{x})$.

The quadratic transform in Theorem 1 suffices to deal with the problem in (11), while its matrix version in Theorem 2 is used when we consider the correlated channel case in Section VI. We then show that the (matrix) quadratic transform leads to an iterative optimization with provable convergence.

Theorem 3 (Convergence Analysis [26]): If $\mathbf{x}$ and $\underline{\boldsymbol{\lambda}}$ are optimized alternatingly in (13) or (15), the value of the original objective function is nondecreasing after each iteration. Furthermore, if the original objective function is differentiable, it converges to a stationary point of (12) or (14).

IV. NONORTHOGONAL PILOT DESIGN

In this section we explore the use of the quadratic transform in nonorthogonal pilot design based on minimizing the MSE. The difficulty of problem (11) lies in its fractional term $\boldsymbol{\phi}_{lk}^H \mathbf{D}_l^{-1} \boldsymbol{\phi}_{lk}$, wherein the numerator and denominator are both affected by the pilot variable $\underline{\boldsymbol{\phi}}$. It is natural to decouple the numerator and denominator by using the quadratic transform in Theorem 1. The resulting problem reformulation is stated in the following proposition.

Proposition 1: The nonorthogonal pilot design problem in (11) is equivalent to

$$\underset{\underline{\boldsymbol{\phi}}, \underline{\boldsymbol{\lambda}}}{\text{maximize}} \quad f(\underline{\boldsymbol{\phi}}, \underline{\boldsymbol{\lambda}}) \qquad (16a)$$

$$\text{subject to} \quad \|\boldsymbol{\phi}_{lk}\|^2 \leq \tau P_{\max} \qquad (16b)$$

$$\boldsymbol{\lambda}_{lk} \in \mathbb{C}^\tau, \qquad (16c)$$

where the new objective function is

$$f(\underline{\boldsymbol{\phi}}, \underline{\boldsymbol{\lambda}}) = \sum_{(l,k)} \alpha_{lk}\Big(2\beta_{llk}\Re\{\boldsymbol{\lambda}_{lk}^H \boldsymbol{\phi}_{lk}\} - \boldsymbol{\lambda}_{lk}^H \mathbf{D}_l \boldsymbol{\lambda}_{lk}\Big). \quad (17)$$

Proof: The reformulation is obtained by treating $\beta_{llk}\boldsymbol{\phi}_{lk}$ and $\mathbf{D}_l$ as $\mathbf{a}_n$ and $\mathbf{B}_n$ in Theorem 1, respectively, along with the nondecreasing function $f_n\big(\mathbf{a}_n^H(\mathbf{x})\mathbf{B}_n^{-1}(\mathbf{x})\mathbf{a}_n(\mathbf{x})\big) = \mathbf{a}_n^H(\mathbf{x})\mathbf{B}_n^{-1}(\mathbf{x})\mathbf{a}_n(\mathbf{x})$. ∎

We propose optimizing $\underline{\boldsymbol{\lambda}}$ and $\underline{\boldsymbol{\phi}}$ alternatingly. As already shown in [7], the auxiliary variable $\underline{\boldsymbol{\lambda}}$ can be optimally updated by solving $\partial f/\partial \boldsymbol{\lambda}_{lk} = \mathbf{0}$ when $\underline{\boldsymbol{\phi}}$ is held fixed, resulting in

$$\boldsymbol{\lambda}_{lk}^\star = \beta_{llk} \mathbf{D}_l^{-1} \boldsymbol{\phi}_{lk}. \qquad (18)$$

It remains to optimize the pilot variable $\underline{\boldsymbol{\phi}}$ for fixed $\underline{\boldsymbol{\lambda}}$. It turns out that the solution can be obtained in closed form. To this

$$R_{lk} = \log_2\left(1 + \frac{\beta_{llk}^2\left(\phi_{lk}^H \mathbf{D}_l^{-1}\phi_{lk}\right)^2}{\frac{1}{M}\left(\sum_{(i,j)}\beta_{lij}+\sigma^2\right)\phi_{lk}^H\mathbf{D}_l^{-1}\phi_{lk} + \sum_{(i,j)}\beta_{lij}^2\phi_{lk}^H\mathbf{D}_l^{-1}\phi_{ij}\phi_{ij}^H\mathbf{D}_l^{-1}\phi_{lk} - \beta_{llk}^2\left(\phi_{lk}^H\mathbf{D}_l^{-1}\phi_{lk}\right)^2}\right). \quad (24)$$

end, we express $f(\underline{\phi}, \underline{\lambda})$ as

$$f(\underline{\phi}, \underline{\lambda}) = \sum_{(l,k)} 2\alpha_{lk}\beta_{llk}\Re\{\lambda_{lk}^H\phi_{lk}\}$$
$$- \sum_{(l,k)}\phi_{lk}^H\left(\sum_{(i,j)}\alpha_{ij}\beta_{ilk}\lambda_{ij}\lambda_{ij}^H\right)\phi_{lk} + \mathrm{const}, \quad (19)$$

in which the last term $\mathrm{const} = \sum_{(l,k)}\alpha_{lk}\sigma^2\|\lambda_{lk}^H\|^2$ does not depend on $\underline{\phi}$. The optimal pilots in terms of $f(\underline{\phi}, \underline{\lambda})$ are then easily solved, resulting in

$$\phi_{lk}^\star = \left(\sum_{(i,j)}\alpha_{ij}\beta_{ilk}\lambda_{ij}\lambda_{ij}^H + \eta_{lk}\mathbf{I}_\tau\right)^{-1}\alpha_{lk}\beta_{llk}\lambda_{lk}, \quad (20)$$

where the Lagrange multiplier η_{lk} accounts for the power constraint and is optimally determined as

$$\eta_{lk}^\star = \begin{cases} 0, \text{ if } \|\phi_{lk}^\star\|^2 \leq \tau P_{\max} \text{ already};\\ \eta_{lk} > 0 \text{ with } \|\phi_{lk}^\star\|^2 = \tau P_{\max}, \text{ otherwise}. \end{cases} \quad (21)$$

The evaluation of (21) can be done by bisection search.

It can be readily obtained from Theorem 3 that the iteration between (18) and (20) leads to convergence.

Proposition 2: The sum of weighted MSEs in (10) is non-increasing after each iteration in Algorithm 1, while the pilot variable $\underline{\phi}$ converges to a stationary point of the nonorthogonal pilot design problem in (11).

To avoid the Lagrange multiplier η_{lk}, we take advantage of the observation in [5] that multiplying all the pilots with the same nonzero scalar δ does not change the MSE values provided that the noise level σ^2 tends to zero. Thus, when the signal-to-noise ratio (SNR) is sufficiently high, we enforce the power constraint by scaling the pilots, without computing the Lagrange multiplier in (21).

Proposition 3 (Nonorthogonal Pilot Design Without Using Lagrange Multiplier): If the noise level $\sigma^2 \to 0$, we can set $\eta_{lk} = 0$ and determine ϕ_{lk} as

$$\phi_{lk}^\star = \delta\tilde{\phi}_{lk}, \text{ for each } (l,k), \quad (22)$$

where $\tilde{\phi}_{lk}$ is obtained from (20) with $\eta_{lk} = 0$ and the scaling factor δ is computed as

$$\delta = \min_{(l,k)}\frac{\sqrt{\tau P_{\max}}}{\|\tilde{\phi}_{lk}\|}. \quad (23)$$

The resulting $\underline{\phi}^\star$ is a stationary point of the nonorthogonal pilot design problem in (11).

Proof: For ease of discussion, we use (P1) to denote the original problem (11), and (P2) the unconstrained version of (11) with the power constraint removed. If $\underline{\phi}'$ is a stationary point of (P2), then it is also a stationary point of (P1) so long as it meets the power constraint automatically.

Algorithm 1: Proposed nonorthogonal pilot design for weighted MSE minimization

---

1 Initialize the pilot variable $\underline{\phi}$ to some feasible value
2 repeat
3   Update the auxiliary variable $\underline{\lambda}$ by (18)
4   *Option 1:* Update the pilots $\underline{\phi}$ by (20) along with the Lagrangian multiplier η_{lk} in (21)
5   *Option 2 (when $\sigma^2 \to 0$):* Update $\underline{\phi}$ by (20) with $\eta_{lk} = 0$, then scale it as in (22) and (23)
6 until *the weighted sum MSE converges*

According to Theorem 3, $\tilde{\phi}$ must be a stationary point of (P2). In addition, it can be shown that the first-order condition of (P2) remains the same after scaling every $\tilde{\phi}_{lk}$ with δ, so $\underline{\phi}^\star$ must be a stationary point of (P2) as well. Note that $\underline{\phi}^\star$ already meets the power constraint because of (23), so it is also a stationary point of (P1). ∎

Algorithm 1 summarizes the main procedure of the proposed nonorthogonal pilot design. Next we discuss the resulting achievable rate. In the existing literature, the achievable rate of massive MIMO systems is typically considered for the orthogonal pilot case. Progress has been made in the recent work [23] to take a special type of nonorthogonal pilots into account. The following theorem provides a further generalization that holds for arbitrary pilots.

Theorem 4 (Achievable Rate with Nonorthogonal Pilots): Given a set of nonorthogonal pilots $\underline{\phi}$ in (5), the data rate R_{lk} in (24) is achievable for user (l, k).

Proof: See Appendix A. ∎

V. ORTHOGONAL PILOT DESIGN

We now assume orthogonal pilots by imposing the constraint (3) on the weighted MMSE problem (11). With each ϕ_{lk} expressed as (p_{lk}, ψ_{lk}), the orthogonal pilot design problem can be formulated as

$$\underset{\underline{p}, \underline{\psi}}{\text{maximize}} \quad \sum_{(l,k)}\alpha_{lk}\beta_{llk}^2 p_{lk}\left(\psi_{lk}^H\mathbf{D}_l^{-1}\psi_{lk}\right) \quad (25a)$$

$$\text{subject to} \quad 0 \leq p_{lk} \leq P_{\max} \quad (25b)$$
$$\psi_{lk} \in \{\varphi_1, \ldots, \varphi_\tau\} \quad (25c)$$
$$\psi_{lk} \neq \psi_{lk'}, \text{ for any } k \neq k', \quad (25d)$$

where the covariance matrix $\mathbf{D}_l$ of $\mathbf{Y}_l$ becomes

$$\mathbf{D}_l = \sigma^2\mathbf{I}_\tau + \sum_{(i,j)}\beta_{lij}p_{ij}\psi_{ij}\psi_{ij}^H. \quad (26)$$

The above problem has a mixed discrete-continuous form since it involves continuous variable $\underline{p}$ and discrete variable $\underline{\psi}$.

Algorithm 2: Proposed orthogonal pilot design for weighted MSE minimization

---

1 Initialize $(p, \underline{\psi})$ to some feasible point

2 repeat

3   Update the auxiliary variable $\underline{\lambda}$ by (18)

4   *Option 1:* Update $(p, \underline{\psi})$ by solving the weighted bipartite matching problem in (33)

5   *Option 2 (when constraint (25d) is removed):* Update $(p, \underline{\psi})$ by the linear search in (36)

6 until *the value of* $\sum_{(l,k)} \alpha_{lk} \mathrm{MSE}_{lk}$ *converges*

A. Orthogonal Pilot Assignment via Power Control

The mixed discrete-continuous problem in (25) is difficult to tackle directly. A naive idea is to reformulate it as a continuous power control problem. Specifically, introducing a new power variable $\tilde{p}_{lk}^{(s)}$ for each user (l, k) and each possible orthogonal pilot φ_s, we optimize the new power variable $\tilde{p}$ and then assign some φ_s with nonzero $\tilde{p}_{lk}^{(s)}$ to each user (l, k).

However, the resulting problem is still difficult because of the orthogonal pilot constraint. Since every user can choose only one pilot in $\{\varphi_1, \ldots, \varphi_\tau\}$, it requires that

$$\left\| \tilde{p}_{lk}^{(1)}, \ldots, \tilde{p}_{lk}^{(\tau)} \right\|_0 = 1, \ l \in [1:L], k \in [1:K]. \quad (27)$$

Moreover, if the users in the same cell cannot choose the same φ_s, then we further have the following constraint:

$$\tilde{p}_{lk}^{(s)} \tilde{p}_{lk'}^{(s)} = 0, \ l \in [1:L], k \neq k' \in [1:K], s \in [1:\tau]. \quad (28)$$

Both (27) and (28) are difficult to handle.

Scalability is another issue since the new power variable $\tilde{p}_{lk}^{(s)}$ needs to be coordinated not only across users, but across all possible pilots φ_s. Thus, rewriting (25) in a continuous form does not necessarily make the problem easier. The rest of this section shows that the mixed discrete-continuous problem in (25) can be efficiently addressed by means of weighted bipartite matching after the quadratic transform.

B. Ratio Decoupling in Orthogonal Pilot Case

The quadratic transform [7] still works in spite of the above changes. Following Proposition 1, we recast problem (25) as

$$\underset{p, \underline{\psi}, \underline{\lambda}}{\text{maximize}} \quad f(p, \underline{\psi}, \underline{\lambda}) \quad (29a)$$

$$\text{subject to} \quad (25b)-(25d) \quad (29b)$$

$$\lambda_{lk} \in \mathbb{C}^\tau, \quad (29c)$$

in which the new objective function is given by

$$f(p, \underline{\psi}, \underline{\lambda}) = \sum_{(l,k)} 2\sqrt{p_{lk}} \alpha_{lk} \beta_{llk} \Re\{\lambda_{lk}^H \psi_{lk}\}$$
$$- \sum_{(l,k)} p_{lk} \psi_{lk}^H \left(\sum_{(i,j)} \alpha_{ij} \beta_{ilk} \lambda_{ij} \lambda_{ij}^H \right) \psi_{lk} + \text{const}, \quad (30)$$

where const refers to terms not depending on $(p, \underline{\psi})$.

As before, we propose to optimize the original variable $(p, \underline{\psi})$ and the auxiliary variable $\underline{\lambda}$ in an iterative fashion.

When $(p, \underline{\psi})$ are held fixed, the optimal $\underline{\lambda}$ is still determined by (18) except that $\underline{\phi}$ is replaced with $(p, \underline{\psi})$. In contrast, the optimization of pilots under fixed $\underline{\lambda}$ is quite different from the nonorthogonal case discussed in the previous section.

The key observation is that due to the convexity of (30), the power variable p_{lk} of user (l, k) can be optimally determined for the new objective function $f(p, \underline{\psi}, \underline{\lambda})$ by solving the first-order equation $\partial f / \partial p_{lk} = 0$, so long as the corresponding normalized sequence ψ_{lk} is fixed. Hence, assuming that $\psi_{lk} = \varphi_s$, for some $s \in [1:\tau]$, the optimal p_{lk} in terms of $f(p, \underline{\psi}, \underline{\lambda})$ can be computed as

$$p_{lk}^{(s)} = \min \left\{ P_{\max}, \left(\frac{\alpha_{lk} \beta_{llk} \Re\{\lambda_{lk}^H \varphi_s\}}{\varphi_s^H \left(\sum_{(i,j)} \alpha_{ij} \beta_{ilk} \lambda_{ij} \lambda_{ij}^H \right) \varphi_s} \right)^2 \right\}. \quad (31)$$

C. Orthogonal Pilot Design via Weighted Bipartite Matching

The new objective function f in (30) plays a crucial role in allowing each p_{lk} to be optimized separately. Otherwise, the optimal p_{lk} in terms of $f(p, \underline{\psi}, \underline{\lambda})$ would depend on the other variables p_{ij} and ψ_{ij} as in the original problem. Given $\psi_{lk} = \varphi_s$, the tentative contribution of user (l, k) to $f(p, \underline{\psi}, \underline{\lambda})$ is

$$\pi_{lk}^{(s)} = 2\sqrt{p_{lk}^{(s)}} \alpha_{lk} \beta_{llk} \Re\{\lambda_{lk}^H \varphi_s\}$$
$$- p_{lk}^{(s)} \varphi_s^H \left(\sum_{(i,j)} \alpha_{ij} \beta_{ilk} \lambda_{ij} \lambda_{ij}^H \right) \varphi_s. \quad (32)$$

As a result, the maximization of $f(p, \underline{\psi}, \underline{\lambda})$ boils down to finding the optimal pair $(\varphi_s, p_{lk}^{(s)})$ for each individual user, recognized as a weighted bipartite matching problem

$$\underset{\underline{x}}{\text{maximize}} \quad \sum_{(l,k,s)} \pi_{lk}^{(s)} x_{lk}^{(s)} \quad (33a)$$

$$\text{subject to} \quad \sum_{s=1}^{\tau} x_{lk}^{(s)} = 1, \text{ for each } (l, k) \quad (33b)$$

$$\sum_{k=1}^{K} x_{lk}^{(s)} \leq 1, \text{ for each } (l, s) \quad (33c)$$

$$x_{lk}^{(s)} \in \{0, 1\}, \quad (33d)$$

where $x_{lk}^{(s)}$ being 1 or 0 indicates whether or not $\psi_{lk} = \varphi_s$, the constraint (33b) implies that each user (l, k) can be assigned only one pilot, and the constraint (33c) implies that the users in the same cell cannot be assigned the same pilot.

The weighted bipartite matching problem in (33) is solvable in polynomial time, e.g., by the Hungarian algorithm [27]. After finding the solution of $\underline{x}$, we recover the solution of the original variables as

$$p_{lk}^\star = \sum_{s=1}^{\tau} x_{lk}^{(s)} p_{lk}^{(s)} \quad \text{and} \quad \psi_{lk}^\star = \sum_{s=1}^{\tau} x_{lk}^{(s)} \varphi_s. \quad (34)$$

The above matching-based optimization is carried out with the auxiliary variable $\underline{\lambda}$ iteratively updated by (18).

Because the orthogonal case involves the discrete variable $\underline{\psi}$, it is hard to establish convergence in terms of $\underline{\psi}$. How-

ever, the convergence of the objective function can still be guaranteed.

Proposition 4: The sum of weighted MSEs in (10) is monotonically decreasing after each iteration in Algorithm 2.

Solving the matching problem in (33) incurs cubic computational complexity $O((K + \tau)^3)$. However, this can be simplified to a linear search if we remove the constraint that the users in the same cell cannot be assigned the same pilot, as specified in the following proposition.

Proposition 5 (Orthogonal Pilot Design via Linear Search): Without the assumption that the users in the same cell cannot be assigned the same pilot, i.e., when constraint (25d) is removed, (p, ψ) can be optimally determined for the new objective function in (33a) as

$$p_{lk}^\star = p_{lk}^{(s_{lk})} \quad \text{and} \quad \psi_{lk}^\star = \varphi_{s_{lk}}, \tag{35}$$

where the index s_{lk} is obtained by the following linear search:

$$s_{lk} = \arg \max_{s \in [1:\tau]} \pi_{lk}^{(s)}. \tag{36}$$

The main steps of the proposed orthogonal pilot design are summarized in Algorithm 2.

D. Max-Min Rate Optimization

So far we have focused on weighted MSE minimization. In this section, we extend the problem setting to rate maximization with max-min fairness as in [11], [23]. Toward this end, we first specialize the rate expression (24) to the orthogonal pilot case in the following corollary which is a well-known result in the literature of massive MIMO.

Corollary 1: As $M \to \infty$, the data rate (24) with orthogonal pilots $p_{lk}\psi_{lk}$ reduces to

$$R_{lk}^\infty = \log_2 \left(1 + \frac{\beta_{llk}^2 p_{lk}^2}{\sum_{(i,j)\neq(l,k)} \beta_{lij}^2 p_{ij}^2 \mathbb{1}_{\psi_{ij}}^{\psi_{lk}}} \right), \tag{37}$$

where the indicator variable $\mathbb{1}_{\psi_{ij}}^{\psi_{lk}}$ equals to 1 if $\psi_{ij} = \psi_{lk}$ and equals to 0 otherwise.

Our goal is to maximize the minimum rate R_{lk}^∞ across all users, i.e.,

$$\underset{p, \psi}{\text{maximize}} \quad \min_{(l,k)} \left\{ R_{lk}^\infty \right\} \tag{38a}$$

$$\text{subject to} \quad (25b)-(25d). \tag{38b}$$

Dropping logarithm and substituting

$$\xi_{lk} = p_{lk}^2, \tag{39}$$

we rewrite (38) as

$$\underset{\xi, \psi}{\text{maximize}} \quad \min_{(l,k)} \left\{ \frac{\beta_{llk}^2 \xi_{lk}}{\sum_{(i,j)\neq(l,k)} \beta_{lij}^2 \xi_{ij} \mathbb{1}_{\psi_{ij}}^{\psi_{lk}}} \right\} \tag{40a}$$

$$\text{subject to} \quad (25b)-(25d). \tag{40b}$$

We propose to optimize ξ and ψ alternatingly. Since the optimization of ψ under fixed ξ, i.e., orthogonal pilot assignment, has been well studied in the existing literature [11], we concentrate on optimizing ξ with ψ held fixed. The key

---

Algorithm 3: Proposed orthogonal pilot design for max-min optimization of data rates

---
1 Initialize (ψ, p) to some feasible point
2 repeat
3   Optimize ψ via the smart pilot assignment [11]
4   **repeat**
5   Update the auxiliary variable $\tilde{\lambda}$ by (41)
6   Update the power variable p by solving the linear program in (33)
7   **until** *the value of λ' converges*
8 until *the value of $\min_{(l,k)} \gamma_{lk}$ converges*

---

step is to recognize (50) as a concave-convex[1] max-min-ratio problem, so the *generalized Dinkelbach's method* [10] can be used to find the optimal solution. Specifically, with an auxiliary variable

$$\lambda' = \min_{(l,k)} \left\{ \frac{\beta_{llk}^2 \xi_{lk}}{\sum_{(i,j)\neq(l,k)} \beta_{lij}^2 \xi_{ij} \mathbb{1}_{\psi_{ij}}^{\psi_{lk}}} \right\}, \tag{41}$$

we decouple the SINRs in (40) as

$$\underset{\xi}{\text{maximize}} \quad \min_{(l,k)} \left\{ \beta_{llk}^2 \xi_{lk} - \sum_{(i,j)\neq(l,k)} \lambda' \beta_{lij}^2 \xi_{ij} \mathbb{1}_{\psi_{ij}}^{\psi_{lk}} \right\} \tag{42a}$$

$$\text{subject to} \quad 0 \leq \xi_{lk} \leq P_{\max}^2. \tag{42b}$$

According to the generalized Dinkelbach's method [10], solving the linear programming problem in (42) iteratively leads to the optimal ξ in (40). Algorithm 3 summarizes the above steps.

We remark that the quadratic transform in Theorem 1 can be applied to the max-min-ratio (40) as well. The corresponding new problem is

$$\underset{\xi}{\text{maximize}} \quad \min_{(l,k)} \left\{ 2\lambda_{lk}\beta_{llk}\sqrt{\xi_{lk}} - \sum_{(i,j)\neq(l,k)} \lambda_{lk}^2 \beta_{lij}^2 \xi_{ij} \mathbb{1}_{\psi_{ij}}^{\psi_{lk}} \right\} \tag{43a}$$

$$\text{subject to} \quad 0 \leq \xi_{lk} \leq P_{\max}^2, \tag{43b}$$

with the auxiliary variable λ_{lk} iteratively updated as

$$\lambda_{lk} = \frac{\beta_{llk}\sqrt{\xi_{lk}}}{\sum_{(i,j)\neq(l,k)} \beta_{lij}^2 \xi_{ij} \mathbb{1}_{\psi_{ij}}^{\psi_{lk}}}. \tag{44}$$

Compared to the quadratic transform, the generalized Dinkelbach's method is more efficient here since it introduces only one auxiliary variable λ'. However, the generalized Dinkelbach's method does not work for a general multi-ratio problem as in (11).

VI. CORRELATED RAYLEIGH FADING

This section aims at an extension of the foregoing algorithmic framework to include channel correlation. We now assume that each Rayleigh fading $\mathbf{g}_{lij}$ is drawn from $\mathcal{CN}(\mathbf{0}, \mathbf{R}_{lij})$ where the covariance matrix $\mathbf{R}_{lij} \in \mathbb{C}^{M \times M}$ is not necessarily

---

[1]An FP problem is said to be concave-convex if its numerator function is concave while its denominator function is convex.

I_M; other settings remain the same as before. The MMSE channel estimate now becomes

$$\hat{\mathbf{h}}_{llk} = \mathbf{W}_{lk}\mathbf{U}_l^{-1}\text{vec}(\mathbf{Y}_l), \qquad (45)$$

where $\mathbf{W}_{lk} \in \mathbb{C}^{M \times \tau M}$ and $\mathbf{U}_{lk} \in \mathbb{C}^{\tau M \times \tau M}$ are given by

$$\mathbf{W}_{lk} = \beta_{llk}\boldsymbol{\phi}_{lk}^H \otimes \mathbf{R}_{llk} \qquad (46)$$

and

$$\mathbf{U}_l = \sigma^2\boldsymbol{I}_{\tau M} + \sum_{(i,j)} \beta_{lij}\boldsymbol{\phi}_{ij}\boldsymbol{\phi}_{ij}^H \otimes \mathbf{R}_{lij}. \qquad (47)$$

The resulting MSE is computed as

$$\text{MSE}_{lk} = \beta_{llk}\text{tr}(\mathbf{R}_{llk}) - \text{tr}\Big(\mathbf{W}_{lk}\mathbf{U}_l^{-1}\mathbf{W}_{lk}^H\Big). \qquad (48)$$

We remark that similar forms of MSE have been derived in [28]–[31], albeit for the single-cell case. The correlated version of problem (11) is therefore

$$\underset{\boldsymbol{\phi}}{\text{maximize}} \quad \sum_{(l,k)} \alpha_{lk}\text{tr}\Big(\mathbf{W}_{lk}\mathbf{U}_l^{-1}\mathbf{W}_{lk}^H\Big). \qquad (49a)$$

$$\text{subject to} \quad \|\boldsymbol{\phi}_{lk}\|^2 \le \tau P_{\max}. \qquad (49b)$$

Observe that $\mathbf{W}_{lk}\mathbf{U}_l^{-1}\mathbf{W}_{lk}^H$ is a matrix ratio. In light of the recently developed matrix FP in [26], our ratio-decoupling approach continues to work for (50), as specified in the following proposition.

Proposition 6: The matrix problem in (49) is equivalent to

$$\underset{\boldsymbol{\phi}, \boldsymbol{\Lambda}}{\text{maximize}} \quad f(\boldsymbol{\phi}, \boldsymbol{\Lambda}). \qquad (50a)$$

$$\text{subject to} \quad \|\boldsymbol{\phi}_{lk}\|^2 \le \tau P_{\max} \qquad (50b)$$

$$\boldsymbol{\Lambda}_{lk} \in \mathbb{C}^{\tau M \times M}, \qquad (50c)$$

where the new objective function is

$$f(\boldsymbol{\phi}, \boldsymbol{\Lambda}) = \sum_{(l,k)} \alpha_{lk}\text{tr}\Big(2\Re\{\mathbf{W}_{lk}\boldsymbol{\Lambda}_{lk}\} - \boldsymbol{\Lambda}_{lk}^H\mathbf{U}_l\boldsymbol{\Lambda}_{lk}\Big). \qquad (51)$$

Proof: The reformulation is obtained by treating $\mathbf{W}_{lk}^H$ as $\mathbf{A}_n(\mathbf{x})$ and $\mathbf{U}_l$ as $\mathbf{B}_n(\mathbf{x})$ in Theorem 2, along with the nondecreasing function $F_n\big(\mathbf{A}_n^H(\mathbf{x})\mathbf{B}_n^{-1}(\mathbf{x})\mathbf{A}_n(\mathbf{x})\big) = \text{tr}\big(\mathbf{A}_n^H(\mathbf{x})\mathbf{B}_n^{-1}(\mathbf{x})\mathbf{A}_n(\mathbf{x})\big)$. ∎

In an iterative fashion, when $\boldsymbol{\phi}$ is fixed, each auxiliary variable $\boldsymbol{\Lambda}_{lk}$ is optimally determined as

$$\boldsymbol{\Lambda}_{lk}^\star = \mathbf{U}_l^{-1}\mathbf{W}_{lk}^H. \qquad (52)$$

This update of $\boldsymbol{\Lambda}$ is optimal regardless of the pilot structure. Before proceeding to the optimization of $\boldsymbol{\phi}$ under fixed $\boldsymbol{\Lambda}$, we introduce some shorthand notation:

- The mth row vector of the matrix $\mathbf{R}_{lij}$ is

$$\mathbf{R}_{lij}^m = (\mathbf{e}_M^m)^\top \mathbf{R}_{lij}. \qquad (53)$$

- The sth $M \times 1$ vector on the mth column of $\boldsymbol{\Lambda}_{lk}$ is

$$\boldsymbol{\Lambda}_{lk}^{m,s} = \Big(\mathbf{E}_{\tau M}^{[1+(s-1)M:sM]}\Big)^\top \boldsymbol{\Lambda}_{lk}\mathbf{e}_M^m. \qquad (54)$$

- The square of $\boldsymbol{\Lambda}_{ij}$ is

$$\widetilde{\boldsymbol{\Lambda}}_{ij} = \boldsymbol{\Lambda}_{ij}\boldsymbol{\Lambda}_{ij}^H. \qquad (55)$$

- The sth $M \times 1$ vector on the $(m+(q+1)M)$th column

of $\widetilde{\boldsymbol{\Lambda}}_{ij}$ is

$$\widetilde{\boldsymbol{\Lambda}}_{ij}^{m,sq} = \Big(\mathbf{E}_{\tau M}^{[1+(s-1)M:sM]}\Big)^\top \widetilde{\boldsymbol{\Lambda}}_{ij}\mathbf{e}_M^{m+(q-1)M}. \qquad (56)$$

Nonorthogonal pilots and orthogonal pilots are discussed separately in what follows.

1) Nonorthogonal Case: In optimizing nonorthogonal pilots, the central idea is to complete the square for each $\boldsymbol{\phi}_{lk}$ in the new objective function $f(\boldsymbol{\phi}, \boldsymbol{\Lambda})$. To this end, we first express $f(\boldsymbol{\phi}, \boldsymbol{\Lambda})$ in an alternative form.

Proposition 7: The objective function $f(\boldsymbol{\phi}, \boldsymbol{\Lambda})$ in (51) can be rewritten as

$$f(\boldsymbol{\phi}, \boldsymbol{\Lambda}) = \sum_{(l,k)} 2\Re\{\boldsymbol{\phi}_{lk}^H\mathbf{v}_{lk}\} - \sum_{(l,k)} \boldsymbol{\phi}_{lk}^H\mathbf{Q}_{lk}\boldsymbol{\phi}_{lk} + \text{const}, \qquad (57)$$

in which const refers to terms not depending on $\boldsymbol{\phi}$, the vector variable $\mathbf{v}_{lk} \in \mathbb{C}^\tau$ is given by

$$\mathbf{v}_{lk} = \sum_{m=1}^M \alpha_{lk}\beta_{llk}\Big(\mathbf{R}_{llk}^m\boldsymbol{\Lambda}_{lk}^{m,1}, \ldots, \mathbf{R}_{llk}^m\boldsymbol{\Lambda}_{lk}^{m,\tau}\Big)^\top, \qquad (58)$$

and the matrix variable $\mathbf{Q}_{lk} \in \mathbb{C}^{\tau \times \tau}$ is defined as

$$\mathbf{Q}_{lk} = \sum_{(i,j,m)} \alpha_{ij}\beta_{ilk}\begin{pmatrix} \mathbf{R}_{ilk}^m\widetilde{\boldsymbol{\Lambda}}_{ij}^{m,11} & \cdots & \mathbf{R}_{ilk}^m\widetilde{\boldsymbol{\Lambda}}_{ij}^{m,1\tau} \\ \vdots & & \vdots \\ \mathbf{R}_{ilk}^m\widetilde{\boldsymbol{\Lambda}}_{ij}^{m,\tau 1} & \cdots & \mathbf{R}_{ilk}^m\widetilde{\boldsymbol{\Lambda}}_{ij}^{m,\tau\tau} \end{pmatrix}. \qquad (59)$$

Proof: See Appendix B. ∎

By completing the square in (57), the optimal $\boldsymbol{\phi}_{lk}$ in terms of $f(\boldsymbol{\phi}, \boldsymbol{\Lambda})$ can be readily obtained as

$$\boldsymbol{\phi}_{lk}^\star = \big(\mathbf{Q}_{lk} + \eta_{lk}\boldsymbol{I}_{\tau M}\big)^{-1}\mathbf{v}_{lk}, \qquad (60)$$

where the Lagrange multiplier η_{lk} is again determined by (21). Furthermore, we can make use of Proposition 3 to simplify the update of $\boldsymbol{\phi}_{lk}$: when the SNR is sufficiently high, we just scale the pilots properly to meet the power constraint, thus getting rid of the Lagrange multiplier η_{lk}.

The convergence of Algorithm 1 as stated in Proposition 2 carries over to this correlated channel case.

2) Orthogonal Case: We next generalize the orthogonal pilot design to correlated Rayleigh fading. The main procedure here follows that of Section V. Replacing $\boldsymbol{\phi}_{lk}$ with $(p_{lk}, \boldsymbol{\psi}_{lk})$ in (57), we express the objective function of the orthogonal pilots in (3) as

$$f(\underline{p}, \boldsymbol{\psi}, \boldsymbol{\Lambda}) = \sum_{(l,k)} 2\sqrt{p_{lk}}\,\Re\{\boldsymbol{\psi}_{lk}^H\mathbf{v}_{lk}\}$$
$$- \sum_{(l,k)} p_{lk}\boldsymbol{\psi}_{lk}^H\mathbf{Q}_{lk}\boldsymbol{\psi}_{lk} + \text{const}, \qquad (61)$$

where const refers to terms not depending on $(\underline{p}, \boldsymbol{\psi})$, and $\mathbf{Q}_{lk}$ is defined in (59).

If a particular normalized pilot $\boldsymbol{\varphi}_s$ is assigned to user (l, k), the corresponding optimal p_{lk} for $f(\underline{p}, \boldsymbol{\psi}, \boldsymbol{\Lambda})$ is given by

$$p_{lk}^{(s)} = \frac{\Re\{\boldsymbol{\varphi}_s^H\mathbf{v}_{lk}\}}{\boldsymbol{\varphi}_s^H\mathbf{Q}_{lk}\boldsymbol{\varphi}_s}. \qquad (62)$$

The contribution of user (l, k) to $f(\underline{p}, \boldsymbol{\psi}, \boldsymbol{\Lambda})$ is then computed

TABLE II
COMPUTATIONAL COMPLEXITY AND COMMUNICATION COMPLEXITY OF PROPOSED ALGORITHMS

	Uncorrelated Channel Case		Correlated Channel Case	
	Computational Complexity	Communication Complexity	Computational Complexity	Communication Complexity
Algorithm 1	$O(K^2L^2\tau^2 + KL\tau^3)$	$O(KL^2\tau)$	$O(K^2L^2M^2\tau^2 + KL\tau^3)$	$O(KL^2M^2\tau)$
Algorithm 2	$O(K^2L^2\tau^3 + K^3L)$	$O(KL^2\tau)$	$O(K^2L^2M^2\tau^3 + K^3L)$	$O(KL^2M^2\tau)$
Algorithm 3	$O(K^{2.5}L^{2.5})$	$O(KL^2\log(\tau))$	–	–

as

$$\pi_{lk}^{(s)} = 2\sqrt{p_{lk}^{(s)}}\Re\{\boldsymbol{\varphi}_s^H \mathbf{v}_{lk}\} - p_{lk}^{(s)}\boldsymbol{\varphi}_s^H \mathbf{Q}_{lk}\boldsymbol{\varphi}_s. \quad (63)$$

We aim to find the optimal assignment of $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_\tau\}$ such that value of $f(\underline{p}, \underline{\boldsymbol{\psi}}, \underline{\boldsymbol{\Lambda}})$ is maximized. This target can be reached by solving the same weighted bipartite matching problem as in (33) except that the link weight is evaluated as (63). Again, if we allow the users in the same cell to be assigned the same pilot, each user (l, k) simply chooses its $(p_{lk}, \boldsymbol{\psi}_{lk})$ according to $\pi_{lk}^{(s)}$ by linear search. The property of Algorithm 2 stated in Proposition 4 continues to hold in the correlated channel case.

In contrast, the max-min-rate method in Algorithm 3 cannot be extended for the correlated channel case because the max-min-ratio problem in (40) is no longer concave-convex.

VII. COMPLEXITY ANALYSIS

This section examines how the computational complexities and communication complexities of the proposed algorithms scale with the system parameters (K, L, M, τ). Our discussion focuses on the general form of each algorithm, e.g., Option 1 of Algorithm 1.

We begin with computational complexities. The following analysis is for each iteration in the proposed algorithms:

- *Algorithm 1:* First, it requires a computational complexity of $O(KL\tau^2)$ to compute each $\mathbf{D}_l$ in (8), then $O(\tau^3)$ operations to compute the inverse matrix $\mathbf{D}_l^{-1}$ in (18); because the above operations are done for each cell l, the overall complexity is $O(KL^2\tau^2 + L\tau^3)$. Second, it requires $O(\tau^2)$ operations to compute each $\boldsymbol{\lambda}_{lk}$ in (18), so that the overall complexity across a total of KL users is $O(KL\tau^2)$. Third, it requires $O(KL\tau^2 + \tau^3)$ operations to compute each $\boldsymbol{\phi}_{lk}$ in (20), leading to the overall computational complexity across all users of $O(K^2L^2\tau^2 + KL\tau^3)$. Summarizing, the overall computational complexity is $O(KL^2\tau^2 + KL\tau^3)$. Likewise, it can be shown that Algorithm 1 for the correlated channel case has a computational complexity of $O(KL^2M^2\tau^2 + KLM^3\tau^3)$.
- *Algorithm 2:* The update of $\underline{\boldsymbol{\lambda}}$ in Algorithm 2 is the same as that in Algorithm 1, with a computational complexity of $O(KL^2\tau^2 + L\tau^3)$. We focus on the matching part. It requires $O(KL\tau^2)$ operations to compute the weight $\pi_{lk}^{(s)}$ with respect to each (l, k, s), leading to a total computational complexity $O(K^2L^2\tau^3)$. The subsequent weighted bipartite matching requires $O((\tau + K)^3)$ operations per cell. Thus, the overall computational complexity of Algorithm 2 is $O(K^2L^2\tau^3) + O((\tau + K)^3 L) = $

$O(K^2L^2\tau^3 + K^3L)$. When extended to the correlated channel case, the algorithm requires a computational complexity of $O(K^2L^2M^2\tau^3 + K^3L)$.

- *Algorithm 3:* First, it requires $O(K^2L^2)$ operations to compute the auxiliary variable λ' in (41). We then update $\underline{p}$ by solving a linear programming problem with KL scalar variables. According to the classic work [32], the problem can be solved with a computational complexity of $O(K^{2.5}L^{2.5})$.

Next we consider the communication complexities. It is assumed that every BS l knows *a priori* the large-scale fading related to its cell, i.e., $\{\beta_{lij}, \forall(i, j)\}$ and $\{\beta_{ilk}, \forall(i, k)\}$, and the correlation matrices $\{\mathbf{R}_{lij}, \forall(i, j)\}$ and $\{\mathbf{R}_{ilk}, \forall(i, k)\}$ in addition for the correlated channel case. Our analysis focuses on the communication of the pilot variable, the power variable, and the auxiliary variable between different cells.

- *Algorithm 1:* Each BS l needs to let other BSs know its pilot variable $\{\boldsymbol{\phi}_{lk} \in \mathbb{C}^\tau, k \in [1 : K]\}$ and auxiliary variable $\{\boldsymbol{\lambda}_{lk} \in \mathbb{C}^\tau, k \in [1 : K]\}$; the resulting total communication complexity across L cells is $O(KL^2\tau)$. In the correlated channel case, the auxiliary variable becomes $\{\boldsymbol{\Lambda}_{lk} \in \mathbb{C}^{\tau M \times M}, k \in [1 : K]\}$, so that the communication complexity rises to $O(KL^2M^2\tau)$.
- *Algorithm 2:* Each BS l needs to let other BSs know the orthogonal pilot indices $s \in [1 : \tau]$ used in its cell, leading to a total communication complexity across L cells of $O(KL^2\log(\tau))$. Likewise, the total cost of communicating the power variable $\underline{p}$ is $O(KL^2)$. The communication of the auxiliary variable $\underline{\boldsymbol{\lambda}}$ has the same complexity as in Algorithm 1. Thus, the overall communication complexity of Algorithm 2 is still $O(KL^2\tau)$. The correlated channel case requires $O(KL^2M^2\tau)$.
- *Algorithm 3:* Differing from Algorithm 1 or 2 that lets each cell update its own variables in a distributed fashion, Algorithm 3 employs a central controller to optimize the entire network. Because the auxiliary variable λ' is computed locally at the central controller, the communication cost is only caused by $\underline{\boldsymbol{\psi}}$ and $\underline{p}$, which amounts to $O(KL\log(\tau))$.

The above complexity results are summarized in Table II.

VIII. NUMERICAL RESULTS

We validate the performance of the proposed algorithms in a wireless network with 7 hexagon-shape cells wrapped around. In each cell, there is one BS located at the center and 6 users uniformly distributed. The BS-to-BS distance equals 500m.
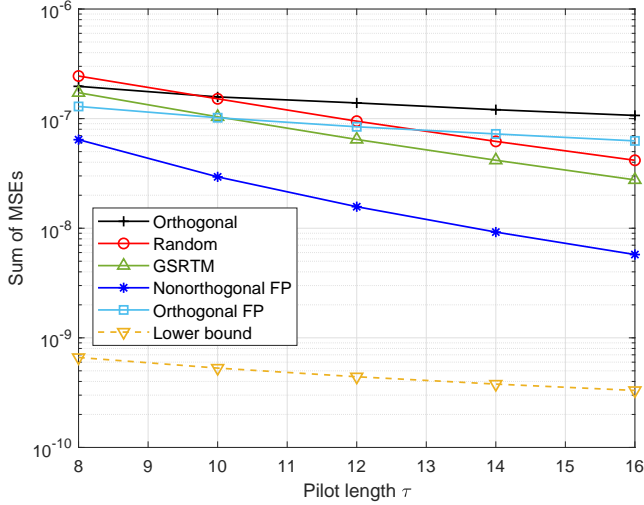
Fig. 2. Sum MSEs minimization in the uncorrelated channel case.

Fig. 3. CDF of MSEs in the uncorrelated channel case.

We assume that each BS has 100 antennas and that a 20MHz-wide spectrum band is reused in each cell. The channel model follows ITU-R M.1225 PedA [33]: the pathloss is computed as $128.1 + 37.6 \log_{10}(d)$ in dB, where d refers to the distance in km; the shadowing in dB between any pair of transmitter and receiver is modeled as an i.i.d. Gaussian random variable drawn from $\mathcal{N}(0, 64)$. Moreover, we set the background noise σ^2 at -169dBm/Hz [34], and set the maximum transmit power level $P_{\max}$ at 23dBm [35].

In the rest of this section, Algorithm 1, Algorithm 2, and Algorithm 3 are referred to as nonorthogonal FP, orthogonal FP, and max-min FP, respectively, and are compared with the following benchmarks:

- *Orthogonal Method:* Fix a set of τ orthogonal pilots, and randomly assign a subset of K orthogonal pilots to each cell; the pilots are all transmitted at full power $P_{\max}$.
- *Random Method:* Generate each pilot symbol randomly and independently according to the Gaussian distribution under the maximum power constraint.
- *GSRTM [5]:* This method optimizes the nonorthogonal pilot symbols successively in order to minimize the sum MSEs of channel estimation.
- *Lower Bound:* We obtain a lower bound on the sum MSEs of channel estimation by ignoring pilot contamination.
- *Smart Pilot Assignment [11]:* It assigns the set of fixed orthogonal pilots to the users in each cell in a greedy fashion iteratively.

We do not compare with the method in [23], because the complexity involved in successive geometric programming only allows very short pilots to be designed.

We first consider minimizing the sum MSEs for uncorrelated channels, i.e., problem (11) with each α_{lk} set to 1. As shown in Fig. 2, the proposed nonorthogonal FP method outperforms the other techniques significantly, e.g., as compared to the random method, it reduces the sum MSEs by around 10dB when $\tau = 16$. The figure shows that the algorithms using nonorthogonal pilots tend to yield lower sum MSEs. But this
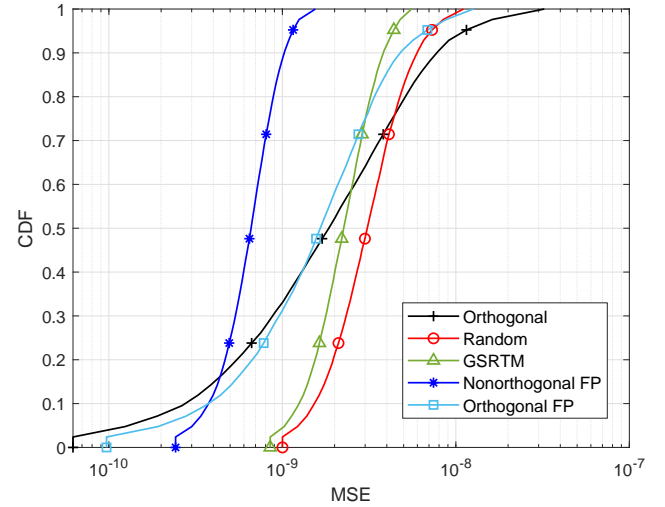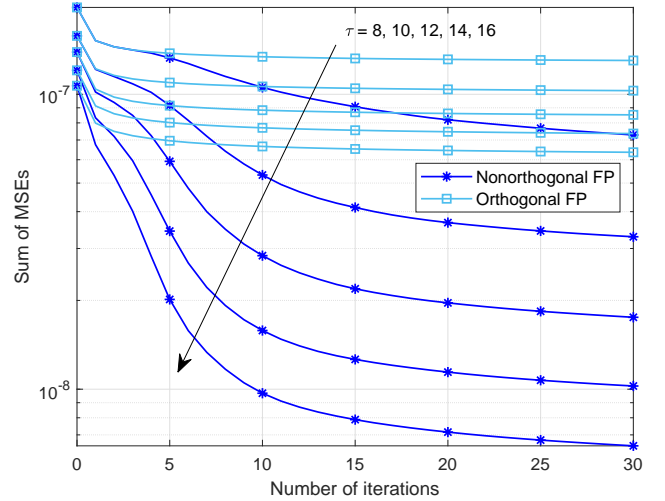

Fig. 4. Convergence of sum MSEs in the uncorrelated channel case.

is not always the case as we observe that the GSRTM method and the random method are inferior to the proposed orthogonal FP approach when τ is sufficiently small. Observe also that the random technique is even worse than the orthogonal method at $\tau = 8$. Thus, nonorthogonal pilot design becomes quite crucial when short pilots are used.

Fig. 3 further shows the cumulative distribution function (CDF) of sum MSEs with respect to a large number of random network realizations. Remarkably, the nonorthogonal FP method achieves smaller sum MSEs than the random and GSRTM techniques at any percentile. The figure also shows that the two orthogonal approaches, the orthogonal FP method and the orthogonal baseline, outperform the nonorthogonal FP algorithm only in the low MSE regime, but are much worse elsewhere.

Fig. 4 shows the convergence of sum MSEs for the proposed FP methods. The two algorithms are both initialized by the
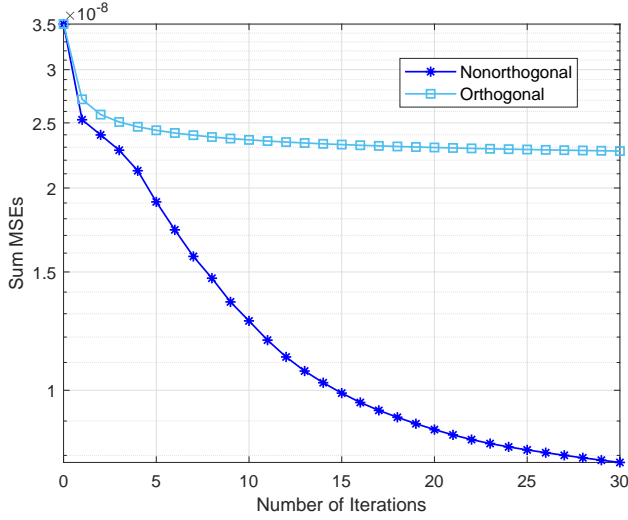
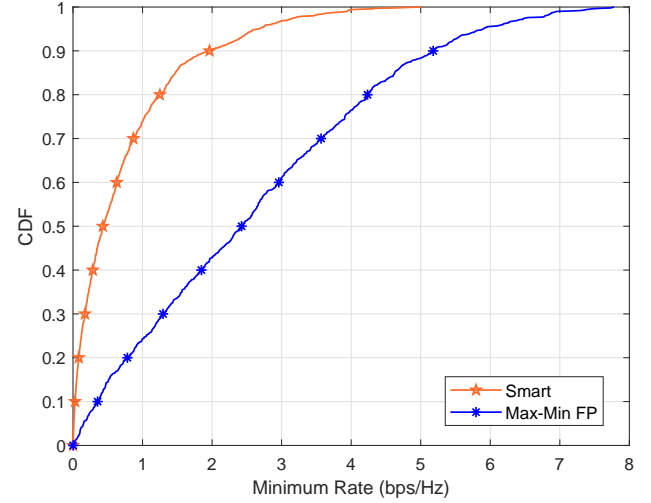Fig. 5. Convergence of sum MSEs in the correlated channel case.

Fig. 7. Max-min-rate optimization in the uncorrelated channel case.
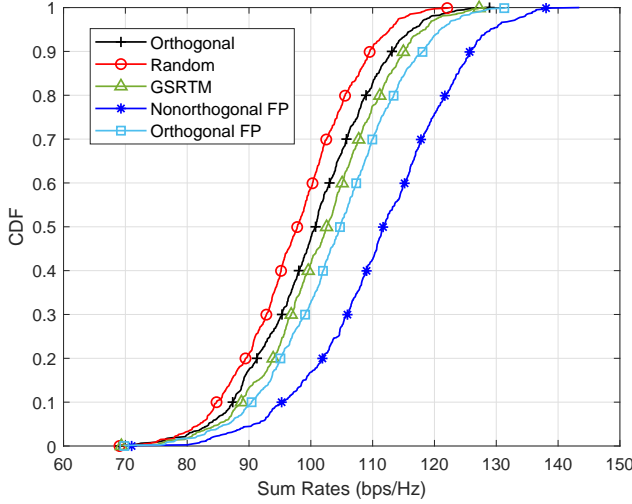

Fig. 6. CDF of sum rates in the uncorrelated channel case.

orthogonal baseline. Observe that the orthogonal FP method converges after only 10 iterations. In comparison, it takes many more iterations for the nonorthogonal FP approach to converge, but the improvement achieved in the first few iterations is considerable, e.g., the sum MSEs is already reduced by 10dB after 10 iterations $\tau = 16$.

We further plot in Fig. 5 the convergence of sum MSEs for a correlated channel case with $\tau = 5$. For ease of simulations, we set $K = 3$, the other parameters remaining the same as before. The channel covariance matrix $\mathbf{R}_{lij}$ is obtained from the exponential model in [23], [36]: randomly generate $\omega_{lij} = \nu e^{j\theta}$ wherein ν is set to 0.5 and θ is drawn i.i.d. from the uniform distribution $U[0, 2\pi)$, then set the (m,n)th entry of the matrix $\mathbf{R}_{lij}$ as

$$\mathbf{R}_{lij}^{m,n} = \begin{cases} \omega_{lij}^{m-n}, & \text{if } m \geq n; \\ (\mathbf{R}_{lij}^{m,n})^H, & \text{otherwise.} \end{cases} \quad (64)$$

The convergence shown in Fig. 5 is slightly slower than that of the uncorrelated channel case, but their overall profiles are similar.

The remainder of this section considers uncorrelated channels along with $K = 6$ and $\tau = 10$. We now compare the sum rate performance of the different methods, assuming that pilots have been optimized for the normalized sum MSEs with $\alpha_{lk} = 1/\beta_{llk}$ [20]–[22], [25]. According to Fig. 6, the nonorthogonal FP algorithm improves upon the other techniques by about 10% at the 50th percentile, even without using any sophisticated power control and receiver design as assumed in (24). It is also worthwhile to note that the orthogonal FP approach is superior to GSRTM, although the latter allows arbitrary nonorthogonal pilots.

Finally, we evaluate the performance of the max-min FP approach. For simplicity, the outer-loop iteration in Algorithm 3 is run only once. Despite this simplification, the max-min FP method can already significantly outperform the smart pilot assignment in [11]. For example, max-min FP reaches a minimum rate higher than 2 Mbps/Hz with probability 58%, whereas the smart pilot assignment only with 10%.

IX. CONCLUSION

This work proposes an FP framework for coordinating the uplink pilots across multiple cells in order to mitigate pilot contamination in massive MIMO. This approach produces a closed-form method for the nonorthogonal pilot design, and a weighted bipartite matching for orthogonal pilot assignment and power control. Further extensions to the max-min-rate optimization and the correlated channel estimation are obtained using max-min-ratio FP and matrix FP, respectively. Numerical results show that the proposed algorithms are superior to the state-of-the-art techniques for both weighted MSE minimization and max-min-rate optimization.

APPENDIX A
PROOF OF THEOREM 4

Let $\tilde{x}_{lk} \sim \mathcal{CN}(0,1)$ be the data signal transmitted from user (l,k) and let $\tilde{\mathbf{z}} \sim \mathcal{CN}(\mathbf{0}, \sigma \boldsymbol{I}_M)$ be the AWGN at BS l during the uplink transmission. With respect to each user (l,k), BS l uses the conjugate transpose of the corresponding channel estimation $\hat{\mathbf{h}}_{llk}$ as the maximum-ratio combining (MRC) receiver, thus obtaining

$$\tilde{y}_{lk} = \hat{\mathbf{h}}_{llk}^H \left(\mathbf{h}_{llk}\tilde{x}_{lk} + \sum_{(i,j)\neq(l,k)} \mathbf{h}_{lij}\tilde{x}_{ij} + \tilde{\mathbf{z}}_l \right). \quad (65)$$

Because the exact value of $\mathbf{h}_{llk}$ is unknown at BS l, it is difficult to compute the achievable rate based on (65) directly. Following the use-and-then-forget method in [37], we introduce the correlation variable

$$\mu_{lk} = \mathbb{E}\left[\hat{\mathbf{h}}_{llk}^H \mathbf{h}_{llk} \right], \quad (66)$$

and rewrite (65) as

$$\tilde{y}_{lk} = \mu_{lk}\tilde{x}_{lk} + \Delta_{lk}, \quad (67)$$

where

$$\Delta_{lk} = \sum_{(i,j)} \hat{\mathbf{h}}_{lij}^H \mathbf{h}_{lij}\tilde{x}_{ij} + \hat{\mathbf{h}}_{llk}^H \tilde{\mathbf{z}}_{lk} - \mu_{lk}\tilde{x}_{lk}. \quad (68)$$

It turns out that $\mathbb{E}[(\mu_{lk}x_{lk})^H \Delta_{ik}] = 0$ (with the expectation taken over all random variables from channel estimation and data transmission, i.e., $\underline{\mathbf{h}}, \underline{\mathbf{Z}}, \underline{\tilde{x}}, \underline{\tilde{\mathbf{z}}}$), so (67) can be treated as if $\tilde{x}_{lk}$ passed through the channel μ with added uncorrelated noise Δ_{lk}.

Most importantly, the exact value of μ_{lk} is available to BS l so long as it knows how the channels and noise are distributed. Henceforth, according to (67), the following data rate is achievable for user (i,k):

$$R_{lk} = \log_2\left(1 + |\mu_{lk}|^2 \cdot \frac{\mathbb{E}[|\tilde{x}_{lk}|^2]}{\mathbb{E}[|\Delta_{lk}|^2]} \right), \quad (69)$$

which can be rewritten as in (24) after some algebra.

APPENDIX B
PROOF OF PROPOSITION 7

We first introduce a lemma used to simplify the calculation with a Kronecker product.

Lemma 1: The following identity holds true given any $\mathbf{a} \in \mathbb{C}^{n_1}$, $\mathbf{b} \in \mathbb{C}^{n_2}$, $\mathbf{C} \in \mathbb{C}^{n_3 \times n_4}$, and $\mathbf{F} \in \mathbb{C}^{n_2 n_4 \times n_1 n_3}$:

$$\text{tr}\left(\left((\mathbf{a}\mathbf{b}^H) \otimes \mathbf{C} \right)\mathbf{F} \right) = \mathbf{b}^H \mathbf{T}\mathbf{a}, \quad (70)$$

where the (i,j)th entry of $\mathbf{T} \in \mathbb{C}^{n_2 \times n_1}$ is computed as

$$T_{ij} = \sum_{m=1}^{n_3} (\mathbf{e}_{n_3}^m)^\top \mathbf{C} \left(\mathbf{E}_{n_3}^{[1+(i-1)n_4:in_4]} \right)^\top \mathbf{F}\, \mathbf{e}_{n_1 n_3}^{j+(m-1)n_1}. \quad (71)$$

Observe that $(\mathbf{e}_{n_3}^m)^\top \mathbf{C}$ corresponds to the mth row of $\mathbf{C}$ while $\left(\mathbf{E}_{n_3}^{[1+(i-1)n_4:in_4]} \right)^\top \mathbf{F}\, \mathbf{e}_{n_1 n_3}^{j+(m-1)n_1}$ corresponds to the ith $n_4 \times 1$ vector on the $(j+(m-1)n_1)$th column of $\mathbf{F}$. The proof is based on expanding the Kronecker product $(\mathbf{a}\mathbf{b}^H) \otimes \mathbf{C}$, followed by some elementary linear algebra.

We now return to the new objective function $f(\underline{\phi}, \underline{\Lambda})$ in (51). Its positive terms can be rewritten as

$$\sum_{(l,k)} \alpha_{lk}\text{tr}\left(2\Re\{\mathbf{W}_{lk}\mathbf{\Lambda}_{lk}\} \right)$$

$$= \sum_{(l,k)} \alpha_{lk}\text{tr}\left(2\Re\{\beta_{llk}\phi_{lk}^H \otimes \mathbf{R}_{llk}\mathbf{\Lambda}_{lk}\} \right) \quad (72a)$$

$$= \sum_{(l,k)} 2\Re\left\{ \text{tr}\left((\alpha_{lk}\beta_{llk}\phi_{lk}^H \otimes \mathbf{R}_{llk})\mathbf{\Lambda}_{lk} \right) \right\} \quad (72b)$$

$$= \sum_{(l,k)} 2\Re\{ \phi_{lk}^H \mathbf{v}_{lk} \}, \quad (72c)$$

where the last equality is due to Lemma 1 with $\mathbf{a}$, $\mathbf{b}$, $\mathbf{C}$, and $\mathbf{F}$ set to $2\alpha_{lk}\beta_{llk}\phi_{lk}^H$, 1, $\mathbf{R}_{llk}$, and $\mathbf{\Lambda}_{lk}$, respectively. Furthermore, the negative terms of $f(\underline{\phi}, \underline{\Lambda})$ can be rewritten as

$$\sum_{(l,k)} \alpha_{lk}\text{tr}\left(\mathbf{\Lambda}_{lk}^H \mathbf{U}_l \mathbf{\Lambda}_{lk} \right)$$

$$= \sum_{(l,k)} \alpha_{lk}\text{tr}\left(\mathbf{U}_l \widetilde{\mathbf{\Lambda}}_{lk} \right) \quad (73a)$$

$$= \sum_{(l,k)} \alpha_{lk}\text{tr}\left(\sum_{(i,j)} \left((\beta_{lij}\phi_{ij}\phi_{ij}^H) \otimes \mathbf{R}_{lij} \right) \widetilde{\mathbf{\Lambda}}_{lk} \right) + \text{const} \quad (73b)$$

$$\overset{(*)}{=} \sum_{(l,k)} \alpha_{lk}\left(\sum_{(i,j)} \beta_{lij}\left(\phi_{ij}^H \mathbf{T}_{lij}\phi_{ij} \right) \right) + \text{const} \quad (73c)$$

$$= \sum_{(l,k)} \phi_{lk}^H \text{tr}\left(\sum_{(i,j)} \alpha_{ij}\beta_{ilk}\mathbf{T}_{ilk} \right)\phi_{lk} + \text{const} \quad (73d)$$

$$= \sum_{(l,k)} \phi_{lk}^H \mathbf{Q}_{lk}\phi_{lk} + \text{const}, \quad (73e)$$

where $\text{const} = \tau\sigma^2 \sum_{(l,k)} \alpha_{lk}\text{tr}(\widetilde{\mathbf{\Lambda}}_{lk})$ does not depend on $\underline{\phi}$; step $(*)$ follows Lemma 1 by letting $\mathbf{a} = \mathbf{b} = \phi_{ij}$, $\mathbf{C} = \mathbf{R}_{lij}$, and $\mathbf{F} = \widetilde{\mathbf{\Lambda}}_{lk}$. Combining (72c) and (73e) gives the new form of $f(\underline{\phi}, \underline{\Lambda})$ in (57).

REFERENCES

[1] K. Shen, Y. C. Eldar, and W. Yu, "Coordinated pilot design for massive MIMO," in *IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, May 2019.

[2] E. G. Larsson, O. Edors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[4] S. S. Ioushua and Y. C. Eldar, "Pilot contamination mitigation with reduced RF chains," in *IEEE Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, July 2017.

[5] S. S. Ioushua and Y. C. Eldar, "Pilot sequence design for mitigating pilot contamination with reduced RF chains," *IEEE Trans. Commun.*, 2020.

[6] 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, May 2018.

[7] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, Mar. 2018.

[8] K. Shen and W. Yu, "Fractional programming for communication systems—Part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, Mar. 2018.

[9] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 133, no. 7, pp. 492–498, Mar. 1967.

[10] J.-P. Crouzeix, "Algorithms for generalized fractional programming," *Mathematical Programming*, vol. 52, no. 1, pp. 191–207, May 1991.

[11] X. Zhu, Z. Wang, L. Dai, and C. Qian, "Smart pilot assignment for massive MIMO," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1644–1647, Sept. 2015.

[12] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[13] H. Q. Ngo and E. G. Larsson, "Blind estimation of effective downlink channel gains in massive MIMO," in *IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP)*, Apr. 2015.

[14] T. Peken, G. Vanhoy, and T. Bose, "Blind channel estimation for massive MIMO," *Analog Integr. Circuits Signal Process.*, vol. 91, pp. 257–266, Feb. 2017.

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

[16] L. Su and C. Yang, "Fractional frequency reuse aided pilot decontamination for massive MIMO systems," in *IEEE Veh. Tech. Conf. (VTC Spring)*, May 2015.

[17] X. Yan, H. Yin, M. Xia, and G. Wei, "Pilot sequences allocation in TDD massive MIMO systems," in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 1488–1493.

[18] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2363–2378, May 2017.

[19] K. Guo, Y. Guo, G. Fodor, and G. Ascheid, "Uplink power control with mmse receiver in multi-cell MU-massive MIMO systems," in *IEEE ICC*, June 2014.

[20] P. Liu, T. Jiang, Q. Zhang, and M. Matthaiou, "Pilot power allocation through user grouping in multi-cell massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1561–1574, Apr. 2017.

[21] J. Zhang, B. Zhang, S. Chen, X. Mu, M. El-Hajjar, and L. Hanzo, "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 759–772, Oct. 2014.

[22] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815–830, Oct. 2014.

[23] 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, Mar. 2018.

[24] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multicell large scale antenna systems," in *IEEE Int. Symp. Inf. Theory (ISIT)*, July 2012, pp. 1137–1141.

[25] T. E. Bogale and L. B. Le, "Pilot optimization and channel estimation for multiuser massive MIMO systems," in *Ann. Conf. Inf. Sci. Sys. (CISS)*, Mar. 2014.

[26] K. Shen, W. Yu, L. Zhao, and D. P. Palomar, "Optimization of MIMO device-to-device networks via matrix fractional programming: A minorization-maximization approach," *IEEE/ACM Trans. Netw.*, vol. 27, no. 5, pp. 2164–2177, Oct. 2019.

[27] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quart.*, vol. 2, no. 1, pp. 83–97, Mar. 1955.

[28] J.H. Kotecha and A.M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[29] Y. Liu, T.F. Wong, and W.W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486–1497, Apr. 2007.

[30] J. Pang, J. Li, Z. Lu, L. Zhao, and L. Chen, "Optimal training sequences for MIMO systems under correlated fading," *J. Syst. Eng. Elect.*, vol. 19, no. 1, pp. 33–38, Feb. 2008.

[31] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.

[32] P. M. Vaidya, "Speeding-up linear programming using fast matrix multiplication," in *IEEE Symp. Found. Comput. Sci. (FOCS)*, 1989.

[33] W. Yu, T. Kwon, and C. Shin, "Multicell coordination via joint scheduling, beamforming and power spectrum adaptation," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 1–14, July 2013.

[34] E. Björnson, M. Matthaiou, A. Pitarokoilis, and M. Debbah, "Distributed massive MIMO in cellular networks: Impact of imperfect hardware and number of oscillators," in *Proc. 23rd Eur. Signal Process. Conf.*, Sept. 2015, pp. 2436–2440.

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

[36] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlated matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sept. 2001.

[37] T. Marzetta, E. G. Larsson, H. Yang, and H. Ngo, *Fundamentals of Massive MIMO*, Cambridge University Press, Cambridge, U.K., 2016.