COORDINATED PILOT DESIGN FOR MASSIVE MIMO

Kaiming Shen[⋆] *Yonina C. Eldar*[†] *Wei Yu*[⋆]

[⋆]Electrial and Computer Enginnering Dept., University of Toronto, Canada
[†]Electrical Engineering Dept., Technion - Israel Institute of Technology, Israel
Emails: kshen@ece.utoronto.ca, yonina@ee.technion.ac.il, weiyu@ece.utoronto.ca

ABSTRACT

Pilot contamination is a main limiting factor in multi-cell massive multiple-input multiple-output (MIMO) systems due to the non-orthogonality of pilot sequences which can seriously impair the channel measurement. Recent work has suggested coordinating pilot sequence design across multiple cells by choosing the sequences to minimize the channel estimation error. This paper further investigates this approach using a new optimization framework. Specifically, we reformulate the weighted minimum mean-squared error (MMSE) measure as a sum-of-functions-of-matrix-ratio program that can be efficiently solved via a matrix fractional programming approach. The proposed algorithm provides fast convergence to a stationary-point solution of the MMSE problem. Simulations demonstrate the advantage of the proposed method over alternative approaches in enhancing channel estimation accuracy.

1. INTRODUCTION

The acquisition of channel state information is crucial in massive multiple-input multiple-output (MIMO) 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 non-orthogonality of the pilots (e.g., due to the reuse of pilots across cells) causes the channel estimate of a particular user terminal to be affected by the pilots of other users—a phenomenon known as *pilot contamination* [**?**, **?**].

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

The works in [**?**, **?**] suggested mitigating pilot contamination by designing the sequences to optimize a suitable system metric. In particular, the authors considered minimizing the sum mean-squared error (MSE) that can be obtained when estimating the channels from the system output. Here we further consider minimizing the weighted sum MSE in channel estimation across a multi-cell massive MIMO network. The main novelty of this paper is in pointing out that this problem is a sum-of-functions-of-matrix-ratio problem that can be efficiently solved by using a recently developed matrix fractional programming (FP) technique [**?**]. In contrast to [**?**] that designs the pilots of all users in a greedy symbols-by-symbols fashion, this work seeks a joint optimization of all pilot symbols based on a matrix FP formulation. Further, we discuss a discrete symbol case in which the pilot symbols are restricted to a fixed constellation. Numerical results show that our coordinated pilot sequence design strategy significantly outperforms both the conventional pilot-reuse strategy and the greedy pilot design technique of [**?**].

The pilot contamination problem has been studied extensively in the literature over the past decade. To quantify the performance, many works [**?**, **?**, **?**, **?**, **?**, **?**, **?**] consider the maximization of the throughput, while many others consider the minimization of the MSE in channel estimation [**?**, **?**, **?**, **?**, **?**, **?**]. The blind or semi-blind pilot methods in [**?**, **?**, **?**] aim to bypass the pilot contamination problem by estimating the channels directly without the pilots. The primary idea of [**?**] is to optimize the set of downlink pilots by using a precoding matrix, while other works, e.g., [**?**, **?**, **?**, **?**, **?**, **?**], seek a contamination-aware allocation of a given set of orthogonal pilots. Furthermore, [**?**, **?**, **?**] consider the optimal pilot design for each cell sequentially.

Notation: We use bold lower-case (or upper-case) letters to denote vectors (or matrices), $\|\cdot\|$ as the Euclidean norm, $(\cdot)^\top$ as the transpose, $(\cdot)^*$ as the conjugate transpose, $\mathrm{tr}(\cdot)$ as the trace, and $(\cdot)^{\frac{1}{2}}$ as the square root of a matrix. Let $\mathbb{E}$ be the expectation, $\mathbb{C}^{m\times n}$ the set of $m\times n$ complex matrices, $\mathbb{H}^{n\times n}$ the set of $n\times n$ positive-definite Hermitian matrices, diag the diagonal matrix, j the imaginary unit, $\Re$ (or $\Im$) the real (or imaginary) part of a complex number, $\boldsymbol{I}_n$ the $n\times n$ identity matrix, and $\mathcal{N}$ (or $\mathcal{CN}$) a (complex) Gaussian distribution. Finally, we use underline to denote a collection of variables, e.g., $\underline{\mathbf{X}} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$.

2. PROBLEM FORMULATION

Consider a total of L cells with one BS and K user terminals per cell. The full spectrum band is reused in every cell. We use i or j to denote the index of each cell and its BS, and (i, k) the index of the kth user in cell i. Assume that every BS has N antennas and every user terminal has a single antenna. Let $\mathbf{h}_{ijk} \in \mathbb{C}^N$ be the channel from user (j, k) to BS i, and let $\mathbf{H}_{ij} = [\mathbf{h}_{ij1}, \mathbf{h}_{ij2}, \ldots, \mathbf{h}_{ijK}]$; each $\mathbf{h}_{ijk}$ is modeled as

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

where $\mathbf{g}_{ijk} \in \mathbb{C}^N$ is small-scale fading with i.i.d. entries distributed as $\mathcal{CN}(0, 1)$, and $\beta_{ijk} \geq 0$ is large-scale fading.

Each BS i estimates its $\mathbf{H}_{ii}$ based on the *uplink* pilot signals from the users in the cell. Let $\mathbf{s}_{ik} \in \mathbb{C}^\tau$ be a sequence of pilot symbols transmitted from user (i, k), and let $\mathbf{S}_i = [\mathbf{s}_{i1}, \mathbf{s}_{i2}, \ldots, \mathbf{s}_{iK}]$. The pilot signal received at BS i is

$$\mathbf{V}_i = \mathbf{H}_{ii}\mathbf{S}_i^\top + \sum\nolimits_{j\neq i} \mathbf{H}_{ij}\mathbf{S}_j^\top + \mathbf{Z}_i \tag{2}$$

where $\mathbf{Z}_i \in \mathbb{C}^{N\times\tau}$ is additive noise with i.i.d. entries distributed as $\mathcal{CN}(0, \sigma^2)$.

Pilot contamination arises if pilot sequences within and across $\mathbf{S}_i$ and $\mathbf{S}_j$ are non-orthogonal. Let $\widehat{\mathbf{H}}_{ii}$ be the *minimum mean square error* (MMSE) estimate of $\mathbf{H}_{ii}$ at BS i, and let $\mathsf{MSE}_{ik} = \mathbb{E}\big[\|\widehat{\mathbf{h}}_{iik} - \mathbf{h}_{iik}\|^2\big]$ be the corresponding MSE for user (i, k), where $\widehat{\mathbf{h}}_{iik}$ is the kth column of $\widehat{\mathbf{H}}_{ii}$. We aim to minimize a weighted sum MSE:

$$\text{minimize} \quad \sum_{i=1}^{L}\sum_{k=1}^{K} w_{ik}\mathsf{MSE}_{ik} \tag{3}$$

given a set of nonnegative weights $w_{ik} \geq 0$. For instance, we may set $w_{ik} = 1$ to minimize the sum MSE as in [**?**], or set $w_{ik} = 1/\beta_{iik}$ to minimize the normalized sum MSE as in [**?**]. Following the steps in [**?**], we can formalize (3) as

$$\underset{\underline{\mathbf{S}}}{\text{maximize}} \quad \sum_{i=1}^{L} \mathrm{tr}\big(\mathbf{W}_i\mathbf{P}_{ii}\mathbf{S}_i^*\mathbf{D}_i^{-1}\mathbf{S}_i\mathbf{P}_{ii}\big) \tag{4a}$$

$$\text{subject to} \quad \|\mathbf{s}_{ik}\|^2 \leq \rho,\ \forall(i, k) \tag{4b}$$

where $\mathbf{W}_i = \mathrm{diag}[w_{i1}, w_{i2}, \ldots, w_{iK}]$, ρ is the power constraint, $\mathbf{P}_{ij} = \mathrm{diag}[\beta_{ij1}, \beta_{ij2}, \ldots, \beta_{ijK}]$, and $\mathbf{D}_i = \sigma^2\boldsymbol{I}_\tau + \sum_{j=1}^{L}\mathbf{S}_j\mathbf{P}_{ij}\mathbf{S}_j^*$.

Problem (4) is a difficult optimization problem, because the choice of pilot sequences $\mathbf{S}_i$ appears in both the numerator and the denominator of a matrix fraction in (4a). While the earlier work [**?**] proposes a greedy sum of ratio traces maximization (GSRTM) algorithm to optimize each row of $\mathbf{S}_i$ sequentially, we devise a matrix-FP approach to optimize the entire matrix $\mathbf{S}_i$ jointly. As a result, our approach has a lower computational complexity, and achieves much higher channel estimation accuracy according to the numerical results.

3. COORDINATED PILOT DESIGN

3.1. Matrix FP

This section reviews the matrix FP technique in [**?**] briefly. We start with the definition of *matrix ratio*. For a pair of matrices $\mathbf{A} \in \mathbb{C}^{m\times n}$ and $\mathbf{B} \in \mathbb{C}^{m\times m}$, $\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}$ is said to be the ratio between the numerator matrix $\mathbf{A}\mathbf{A}^*$ and the denominator matrix $\mathbf{B}$. The following theorem from [**?**] is able to decouple the numerator and denominator of the matrix ratio.

Theorem 1 (Matrix Quadratic Transform [**?**])**.** *Given a nonempty constraint set $\mathcal{X}$ as well as a sequence of functions $\mathbf{A}_i(\mathbf{x}) \in \mathbb{C}^{m\times n}$, functions $\mathbf{B}_i(\mathbf{x}) \in \mathbb{H}^{m\times m}$, and nondecreasing functions $f_i(\mathbf{M})$ in the sense that $f_i(\mathbf{M}') \geq f_i(\mathbf{M})$ if $\mathbf{M}' \succeq \mathbf{M}$, for $i = 1, 2, \ldots, L$, the sum-of-functions-of-matrix-ratio problem*

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \sum_{i=1}^{L} f_i\Big(\mathbf{A}_i^*(\mathbf{x})\mathbf{B}_i^{-1}(\mathbf{x})\mathbf{A}_i(\mathbf{x})\Big) \tag{5}$$

is equivalent to

$$\underset{\mathbf{x}\in\mathcal{X},\,\underline{\mathbf{Y}}}{\text{maximize}} \quad \sum_{i=1}^{L} f_i\Big(2\Re\{\mathbf{A}_i^*(\mathbf{x})\mathbf{Y}_i\} - \mathbf{Y}_i^*\mathbf{B}_i(\mathbf{x})\mathbf{Y}_i\Big) \tag{6}$$

where $\mathbf{Y}_i \in \mathbb{C}^{m\times n}$ is an auxiliary variable introduced for each matrix ratio term.

Proof. By completing the square, each $\mathbf{Y}_i$ can be optimally computed as $\mathbf{Y}_i = \mathbf{B}_i^{-1}(\mathbf{x})\mathbf{A}_i(\mathbf{x})$. Plugging this optimal $\mathbf{Y}_i$ in (6) recovers (5) and thus establishes the equivalence. □

3.2. Iterative Optimization via Matrix FP Approach

In light of Theorem 1, we can reformulate (4) as follows.

Proposition 1. *Problem (4) is equivalent to*

$$\underset{\underline{\mathbf{S}},\underline{\mathbf{Y}}}{\text{maximize}} \quad \sum_{i=1}^{L} \mathrm{tr}\Big(\mathbf{W}_i\big(2\Re\{\mathbf{P}_{ii}\mathbf{S}_i^*\mathbf{Y}_i\} - \mathbf{Y}_i^*\mathbf{D}_i\mathbf{Y}_i\big)\Big) \tag{7a}$$

$$\text{subject to} \quad \|\mathbf{s}_{ik}\|^2 \leq \rho, \tag{7b}$$

$$\mathbf{Y}_i \in \mathbb{C}^{\tau\times K}. \tag{7c}$$

Proof. Note that $\mathbf{P}_{ii} = \mathbf{P}_{ii}^*$. The new objective function is derived by treating $\mathbf{S}_i\mathbf{P}_{ii}^*$ as $\mathbf{A}_i$, $\mathbf{D}_i$ as $\mathbf{B}_i$ in Theorem 1. □

To solve the problem in Proposition 1, we propose to optimize $\underline{\mathbf{S}}$ and $\underline{\mathbf{Y}}$ alternatively. When $\underline{\mathbf{S}}$ is held fixed, each $\mathbf{Y}_i$ can be optimally determined by completing the square for $\mathbf{Y}_i$ in (7a), i.e.,

$$\mathbf{Y}_i = \mathbf{D}_i^{-1}\mathbf{S}_i\mathbf{P}_{ii}. \tag{8}$$

Next, we optimize $\underline{\mathbf{S}}$ for fixed $\underline{\mathbf{Y}}$. The key step is to rewrite the objective function in (7a) as $\sum_{(i,k)}\xi_{ik} + c(\underline{\mathbf{Y}})$ after some

Algorithm 1: Coordinated Pilot Design

Initialize all the variables to feasible values;
repeat
  Update the auxiliary variable $\underline{\mathbf{Y}}$ by (8);
  Update the pilot variable $\underline{\mathbf{S}}$ by (11);
until *the variables* $(\underline{\mathbf{S}}, \underline{\mathbf{Y}})$ *converge*;

matrix algebra, where $\mathbf{y}_{ik}$ refers to the kth column of $\mathbf{Y}_i$,

$$\xi_{ik} = 2\Re\{w_{ik}\beta_{iik}\mathbf{s}_{ik}^*\mathbf{y}_{ik}\} - \mathbf{s}_{ik}^*\left(\sum_{j=1}^{L}\beta_{jik}\mathbf{Y}_j\mathbf{W}_j\mathbf{Y}_j^*\right)\mathbf{s}_{ik}, \tag{9}$$

and $c(\underline{\mathbf{Y}})$ refers to the term that is independent of $\underline{\mathbf{S}}$. Further, with the power constraint (7b) integrated in, we arrive at the following new objective function for problem (7):

$$f(\underline{\mathbf{S}}, \underline{\mathbf{Y}}) = \sum\nolimits_{(i,k)}\left(\xi_{ik} - \lambda_{ik}\left(\|\mathbf{s}_{ik}\|^2 - \rho\right)\right) + c(\underline{\mathbf{Y}}) \tag{10}$$

where each λ_{ik} is a Lagrangian multiplier for constraint (7b). By completing the square in (10), we find the optimal $\mathbf{s}_{ik}$ as

$$\mathbf{s}_{ik} = \left(\sum_{j=1}^{L}\beta_{jik}\mathbf{Y}_j\mathbf{W}_j\mathbf{Y}_j^* + \lambda_{ik}\boldsymbol{I}_\tau\right)^{-1} w_{ik}\beta_{iik}\mathbf{y}_{ik} \tag{11}$$

where λ_{ik} can be computed by a bisection search to meet constraint (7b). Algorithm 1 summarizes the overall approach.

In the particular case where $\mathbf{Z}_i$ is negligible (i.e., $\sigma^2 = 0$), as pointed out in [**?**], the weighted sum MSE remains the same if each pilot $\mathbf{s}_{ik}$ is multiplied by the same nonzero factor α. In this case, we can enforce the power constraint by scaling all the $\mathbf{s}_{ik}$'s (for $\lambda_{ik} = 0$) simultaneously with a sufficiently small $\alpha > 0$, instead of going through the computation of λ_{ik}.

Proposition 2. *The weighted sum MSE is monotonically non-increasing after each iterate of Algorithm 1. Further, the pilot variable* $\underline{\mathbf{S}}$ *converges to a stationary point of problem (4).*

Proof. It can be shown that the iterative update by Algorithm 1 can be interpreted as a sequence of minorization-maximization [**?**, **?**]; the technical details follow [**?**]. ☐

We next compare Algorithm 1 with the GSRTM algorithm proposed in [**?**]. The main idea behind GSRTM is to optimize one row of the matrix $\mathbf{S}_i$ at a time while fixing all other rows. Because the rows of $\mathbf{S}_i$ are not optimized jointly in GSRTM, this greedy method is prone to being trapped in a local optimum. Furthermore, it can be shown that Algorithm 1 has a computational complexity scaling of $O(\tau^3 LKU)$ where U is the number of iterations, while GSRTM has computational complexity scaling of $O(\tau^3 L^2 K + \tau^4 L)$, so that GSRTM is more sensitive[1] to τ and L. Moreover, the convergence property of GSRTM is difficult to analyze, whereas Algorithm 1 is guaranteed to converge to a stationary-point solution. In Section 4 we present numerical results that demonstrate the performance advantage of Algorithm 1 over GSRTM.

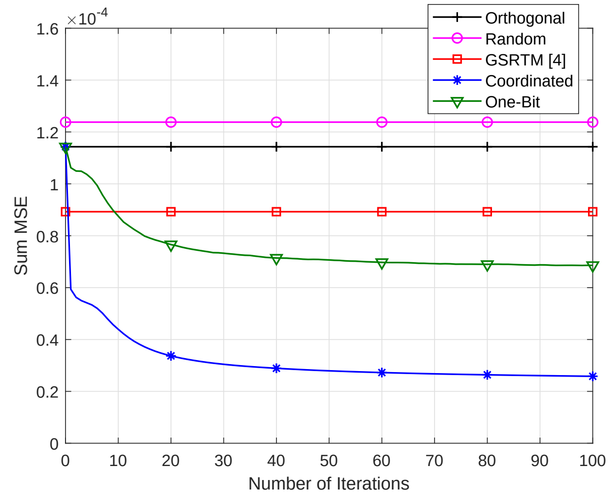

Fig. 1. Sum MSE after each iteration.

3.3. One-Bit Pilot Sequence Design

Arbitrary complex-valued pilot sequences may be difficult to implement in practice. In this section, we restrict the choice of each pilot symbol $s_{ik}[t]$ to a 4-QAM constellation $\mathcal{Q} = \{\varepsilon(1+\mathrm{j}), \varepsilon(1-\mathrm{j}), \varepsilon(-1+\mathrm{j}), \varepsilon(-1-\mathrm{j})\}$ where $\varepsilon = \sqrt{\rho/2\tau}$. Such sequences are referred to as one-bit pilot sequences, because their in-phase and quadrature components are ± 1.

To design optimal one-bit pilot sequences, we maximize the objective function $f(\underline{\mathbf{S}}, \underline{\mathbf{Y}})$ in (10) for fixed $\underline{\mathbf{Y}}$ over the QAM-constellation as follows:

$$\mathbf{s}_{ik} = \arg\min_{\mathbf{q}\in\mathcal{Q}^\tau}\left\|\left(\sum_{j=1}^{L}\beta_{jik}\mathbf{Y}_j\mathbf{W}_j\mathbf{Y}_j^*\right)^{\frac{1}{2}}(\mathbf{q} - \tilde{\mathbf{s}}_{ik})\right\| \tag{12}$$

where $\tilde{\mathbf{s}}_{ik}$ has the same form as (11) but with $\lambda_{ik} = 0$. However, the above projection of $\tilde{\mathbf{s}}_{ik}$ to $\mathcal{Q}^\tau$ may be computationally complex in practice as the size of $\mathcal{Q}^\tau$ grows exponentially with the pilot length τ. Thus, we propose a suboptimal solution of simply rounding each $\tilde{s}_{ik}[t]$ to $\mathcal{Q}$, i.e.,

$$s_{ik}[t] = \varepsilon\cdot\mathrm{sgn}(\Re\{\tilde{s}_{ik}[t]\}) + \mathrm{j}\,\varepsilon\cdot\mathrm{sgn}(\Im\{\tilde{s}_{ik}[t]\}) \tag{13}$$

where $\mathrm{sgn}(\cdot)$ is the sign function. Observe that the heuristic in (13) is equivalent to $\mathbf{s}_{ik} = \arg\min_{\mathbf{q}\in\mathcal{Q}^\tau}\|\mathbf{q} - \tilde{\mathbf{s}}_{ik}\|$. As compared to (12), this heuristic in essence approximates $\left(\sum_{j=1}^{L}\beta_{jik}\mathbf{Y}_j\mathbf{W}_j\mathbf{Y}_j^*\right)^{\frac{1}{2}}$ as $\delta\boldsymbol{I}_\tau$ for some δ.

4. NUMERICAL RESULTS

We validate the performance of the proposed method in a 7-cell wrapped-around network. Each cell consists of a 16-

[1] Although U may increase with τ and L, we can stop Algorithm 1 early because of its monotonic improvement as stated in Proposition 2.

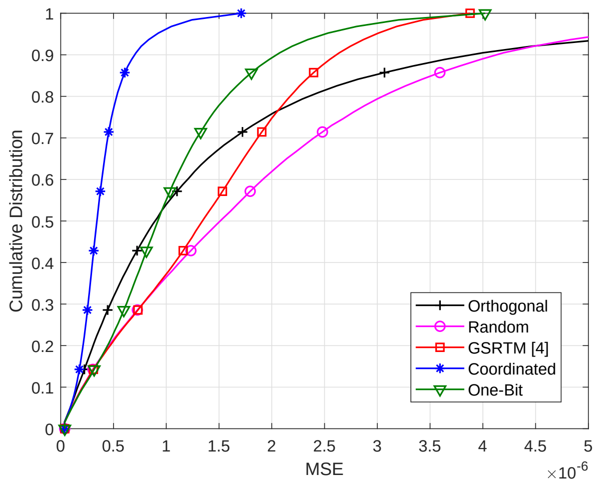

Fig. 2. Cumulative distribution of MSEs.

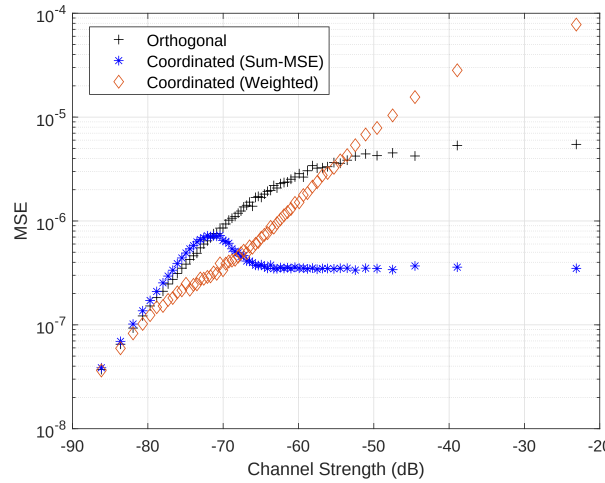

Fig. 3. MSE vs. Large-scale channel strength β_{iik}.

antenna BS located at the center and 9 single-antenna user terminals uniformly distributed. The BS-to-BS distance is 1000 meters. Let $\tau = 10$ and let $\rho = 1$. Following [**?**], we assume that the background noise is negligible and that $\beta_{ijk} = \varphi_{ijk}/d^3_{ijk}$ where φ_{ijk} is an i.i.d. log-normal random variable according to $\mathcal{N}(0, 8^2)$ and d_{ijk} is the distance between user (j, k) and BS i. In addition to the GSRTM algorithm with a random dictionary (see [**?**]), we further introduce two baseline methods as follows:

- *Orthogonal Method:* Fix a set of 10 orthogonal pilots; allocate a random subset of 9 pilots to users in each cell.
- *Random Method:* Generate the pilots randomly and independently according to the Gaussian distribution.

The orthogonal method is used to initialize Algorithm 1.

Fig. 1 compares the sum MSE for the various methods. According to the figure, the coordinated approach in Algorithm 1 reduces the sum MSE sharply as compared to the conventional orthogonal method. Further, around 75% of the sum-MSE reduction is obtained after just 10 iterations. It can also be seen that the one-bit strategy already improves upon the baseline methods and GSRTM, albeit not by as much as the infinite precision coordinated approach. Fig. 2 takes a closer look at the cumulative distribution of the MSE. Observe that the coordinated approach is far superior to all the other techniques in that it yields the smallest MSE in all the percentiles.

We now consider minimizing the weighted sum MSE throughout the network. Because the absolute value of MSE is proportional to the channel magnitude, weighting MSEs equally would give preference to the users with strong channels. To provide some measure of fairness, a possible heuristic [**?**] is to weight the MSEs by $w_{ik} = 1/\beta_{iik}$.

Fig. 3 shows a scatter plot of MSE vs. channel strength for this weighted coordinated approach as compared to the sum-MSE version of Algorithm 1 and the orthogonal method. Although the sum-MSE coordinated approach has considerable advantage over the orthogonal method in minimizing the overall MSEs as shown in the previous results, its performance in the weak-channel region (e.g., when $\beta_{iik} < -70$dB) is close to or even slightly worse than that of the orthogonal method as shown in Fig. 3. The reason is that using the sum MSE as the objective does not take into account the difference in channel strengths among the users, while the weighted coordinated approach is able to improve the MSE for the cell-edge users (which are more vulnerable to pilot contamination) at slight cost to the cell-center users. Indeed, the weighted coordinated approach is inferior to the orthogonal method when the channel strength is very strong ($\beta_{iik} > -55$dB), but only a very small portion of users have such strong channels. Thus, there is overall benefit for the weighted coordinated approach.

5. CONCLUSION

This paper advocates a matrix-FP approach for coordinating the uplink pilots across multiple cells in order to mitigate pilot contamination in a massive MIMO system. The proposed algorithm optimizes the pilots iteratively in closed form, guaranteeing a monotonic reduction of the weighted sum MSE of the channel estimation throughout the network. Numerical results show that the proposed algorithm outperforms the conventional orthogonal pilot reuse method significantly particularly for the cell-edge users and also improves upon a recently proposed greedy method.

6. ACKNOWLEDGEMENT

The authors would like to thank Shahar Stein Ioushua and Dr. Foad Sohrabi for their help in the simulation part of this work.

7. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] S. S. Ioushua and Y. C. Eldar, "Pilot contamination mitigation with reduced RF chains," [Online]. Available: https://arxiv.org/abs/1801.05483, 2018.

[5] K. Shen, W. Yu, L. Zhao, and D. P. Palomar, "Coordinated scheduling and spectrum sharing via matrix fractional programming," [Online]. Available: https://arxiv.org/abs/1808.05678, 2018.

[6] 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.

[7] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multicell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[8] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Intercell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, Feb. 2013.

[9] 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.

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

[11] 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.

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

[13] D. Hu, L. He, and X. Wang, "Semi-blind pilot decontamination for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 525–536, Jan. 2016.

[14] 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.

[15] 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.

[16] 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.

[17] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *IEEE Int. Conf. Acoustics Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3249–3252.

[18] 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.

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

[20] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.