COORDINATED PILOT DESIGN FOR MASSIVE MIMO

Kaiming Shen*  Yonina C. Eldar†  Wei Yu*

*Electrical and Computer Engineering 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 (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), ||·|| as the Euclidean norm, (·)T as the transpose, (·)* as the conjugate transpose, tr(·) as the trace, and (·)1/2 as the square root of a matrix. Let E be the expectation, Cm×n the set of m×n complex matrices, Hn×n the set of n×n positive-definite Hermitian matrices, diag the diagonal matrix, j the imaginary unit, R (or ℍ) the real (or imaginary) part of a complex number, In the n×n identity matrix, and N (or CN) a (complex) Gaussian distribution. Finally, we use underline to denote a collection of variables, e.g., X = {X1, X2, . . . , Xn}.
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 $k$th user in cell $i$. Assume that every BS has $N$ antennas and every user terminal has a single antenna. Let $h_{ijk} \in \mathbb{C}^N$ be the channel from user $(j, k)$ to BS $i$, and let $\mathbf{H}_{ij} = [h_{i,j1}, h_{i,j2}, \ldots, h_{i,jK}]$; each $h_{ijk}$ is modeled as

$$h_{ijk} = g_{ijk} \sqrt{\beta_{ijk}}$$

(1)

where $g_{ijk} \in \mathbb{C}$ 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 $s_{ik} \in \mathbb{C}^T$ be a sequence of pilot symbols transmitted from user $(i, k)$, and let $\mathbf{S}_i = [s_{i1}, s_{i2}, \ldots, s_{iK}]$. The pilot signal received at BS $i$ is

$$\mathbf{v}_i = \mathbf{H}_{ii} \mathbf{S}_i^T + \sum_{j \neq i} \mathbf{H}_{ij} \mathbf{S}_j^T + \mathbf{Z}_i$$

(2)

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

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

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

(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 [2], or set $w_{ik} = 1/\beta_{ik}$ to minimize the normalized sum MSE as in [2]. Following the steps in [2], we can formalize (3) as

$$\text{maximize} \sum_{i=1}^{L} \text{tr}\left(\mathbf{W}_i \mathbf{P}_{ii} \mathbf{S}_i^T \mathbf{D}^{-1}_{ii} \mathbf{S}_i \mathbf{P}_{ii}\right)$$

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

(4a)

where $\mathbf{W}_i = \text{diag}[w_{i1}, w_{i2}, \ldots, w_{iK}]$, $\rho$ is the power constraint, $\mathbf{P}_{ij} = \text{diag}[\beta_{ij1}, \beta_{ij2}, \ldots, \beta_{ijk}]$, and $\mathbf{D}_i = \sigma^2 \mathbf{I}_T + \sum_{j=1}^{L} \mathbf{S}_j \mathbf{P}_{ij} \mathbf{S}_j^T$.

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 [4] 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 [2] briefly. We start with the definition of matrix ratio. For a pair of matrices $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $\mathbf{B} \in \mathbb{C}^{m \times n}$, $\mathbf{A}^* \mathbf{B}^{-1}$ is said to be the ratio between the numerator matrix $\mathbf{A} \mathbf{A}^*$ and the denominator matrix $\mathbf{B}$. The following theorem from [2] is able to decouple the numerator and denominator of the matrix ratio.

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

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

(5)

is equivalent to

$$\text{maximize} \sum_{x \in \mathcal{X}} \sum_{i=1}^{L} f_i(2\Re\{\mathbf{A}_i^*(x) \mathbf{B}_i^{-1}(x) \mathbf{A}_i(x)\})$$

(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 $\mathbf{Y}_i = \mathbf{B}_i^{-1}(x) \mathbf{A}_i(x)$. Plugging this optimal $\mathbf{Y}_i$ in (6) recovers (5) and thus establishes the equivalence. \hfill $\square$

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

$$\text{maximize} \sum_{x \in \mathcal{X}} \sum_{i=1}^{L} \text{tr}\left(\mathbf{W}_i (2\Re\{\mathbf{P}_{ii} \mathbf{S}_i^T \mathbf{Y}_i\} - \mathbf{Y}_i \mathbf{D}_i \mathbf{Y}_i)\right)$$

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

$$\mathbf{Y}_i \in \mathbb{C}^{T \times K}.$$ 

(7a)

(7b)

(7c)

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

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

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

(8)

Next, we optimize $\mathbf{S}$ for fixed $\mathbf{Y}$. The key step is to rewrite the objective function in (7a) as $\sum_{(i,k)} \xi_{ik} + c(\mathbf{Y})$ after some
matrix algebra, where \( y_{ik} \) refers to the \( k \)th column of \( Y_i \),
\[
    \xi_{ik} = 2 \Re \{ w_{ik} \beta_{ik} s_{ik}^* y_{ik} \} - s_{ik}^* \left( \sum_{j=1}^L \beta_{jik} Y_j W_j Y_j^* \right) s_{ik},
\]
and \( c(Y) \) refers to the term that is independent of \( S \). Further, with the power constraint (7b) integrated in, we arrive at the following new objective function for problem (7):
\[
    f(S, Y) = \sum_{(i,k)} \left( \xi_{ik} - \lambda_{ik} (||s_{ik}||^2 - \rho) \right) + c(Y) \tag{10}
\]
where each \( \lambda_{ik} \) is a Lagrangian multiplier for constraint (7b). By completing the square in (10), we find the optimal \( s_{ik} \) as
\[
    s_{ik} = \left( \sum_{j=1}^L \beta_{jik} Y_j W_j Y_j^* + \lambda_{ik} I_r \right)^{-1} w_{ik} \beta_{ik} Y_{ik} \tag{11}
\]
where \( \lambda_{ik} \) can be computed by a bisection search to meet constraint (7b). Algorithm 1 summarizes the overall approach.

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

**Proposition 2.** The weighted sum MSE is monotonically non-increasing after each iterate of Algorithm 1. Further, the pilot variable \( S \) 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 \( S_i \) at a time while fixing all other rows. Because the rows of \( S_i \) 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 \( O(\tau^3 L K U) \) 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 \( \tau \) and \( L \). Moreover, the convergence property of GSRTM is difficult to analyze, whereas Algorithm 1

\(^1\) Although \( U \) may increase with \( \tau \) and \( L \), we can stop Algorithm 1 early because of its monotonic improvement as stated in Proposition 2.

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.

### 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 \( Q = \{ \varepsilon (1 + j), \varepsilon (1 - j), \varepsilon (-1 + j), \varepsilon (-1 - j) \} \) where \( \varepsilon = \sqrt{\rho/2\tau} \). Such sequences are referred to as one-bit pilot sequences, because their in-phase and quadrature components are \( \pm 1 \).

To design optimal one-bit pilot sequences, we maximize the objective function \( f(S, Y) \) in (10) for fixed \( Y \) over the QAM-constellation as follows:
\[
    s_{ik} = \arg \min_{q \in Q^*} \left\| \left( \sum_{j=1}^L \beta_{jik} Y_j W_j Y_j^* \right)^{1/2} (q - \tilde{s}_{ik}) \right\| \tag{12}
\]
where \( \tilde{s}_{ik} \) has the same form as (11) but with \( \lambda_{ik} = 0 \). However, the above projection of \( \tilde{s}_{ik} \) to \( Q^* \) may be computationally complex in practice as the size of \( Q^* \) grows exponentially with the pilot length \( \tau \). Thus, we propose a suboptimal solution of simply rounding each \( \tilde{s}_{ik}[t] \) to \( Q \), i.e.,
\[
    s_{ik}[t] = \varepsilon \cdot \sgn(\Re \{ \tilde{s}_{ik}[t] \}) + j \varepsilon \cdot \sgn(\Im \{ \tilde{s}_{ik}[t] \}) \tag{13}
\]
where \( \sgn(\cdot) \) is the sign function. Observe that the heuristic in (13) is equivalent to \( s_{ik} = \arg \min_{q \in Q^*} \| q - \tilde{s}_{ik} \| \).

As compared to (12), this heuristic in essence approximates \( \left( \sum_{j=1}^L \beta_{jik} Y_j W_j Y_j^* \right)^{1/2} \) as \( \delta I_r \) for some \( \delta \).

### 4. NUMERICAL RESULTS

We validate the performance of the proposed method in a 7-cell wrapped-around network. Each cell consists of a 16-
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 $\beta_{ijk} = \varphi_{ijk}/d_{ijk}^3$ where $\varphi_{ijk}$ is an i.i.d. log-normal random variable according to $\mathcal{N}(0, \sigma^2)$ 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 MSEx 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 $w_{ik} = 1/\beta_{ik}$.

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 $\beta_{ik} < -70$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 ($\beta_{ik} > -55$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


