Massive Multiple-Input-Multiple-Output (M-MIMO) system is a promising technology that offers mobile networks substantial increase in throughput. In Time-Division Duplexing (TDD), the uplink training allows a Base Station (BS) to acquire Channel State Information (CSI) for both uplink reception and downlink transmission. This is essential for M-MIMO systems where downlink training pilots would consume large portion of the bandwidth. In densely populated areas, pilot symbols are reused among neighboring cells. Pilot contamination is the fundamental bottleneck on the performance of M-MIMO systems. Pilot contamination effect in antenna arrays can be mitigated by treating the channel estimation problem in angular domain where channel sparsity can be exploited. In this paper, we introduce a codebook that projects the channel into orthogonal beams and apply Minimum Mean-Squared Error (MMSE) criterion to estimate the channel. We also propose data-aided channel covariance matrix estimation algorithm for angular domain MMSE channel estimator by exploiting properties of linear antenna array. The algorithm is based on simple linear operations and no matrix inversion is involved. Numerical results show that the algorithm performs well in mitigating pilot contamination where the desired channel and other interfering channels span overlapping angle-of-arrivals.

1. Introduction

Catering to throughput/data-rate demands of users in very densely populated areas is costly using legacy solutions. This would typically require operators to have very dense mobile networks cell sites with increase in cost of backhauling, powering, maintaining, and securing the sites. This is particularly critical in emerging markets which will increasingly have the most densely populated areas [1] but low Average Revenues Per Users (ARPUs) [2]. The extreme Mobile BroadBand (eMBB) capabilities envisioned in 5G provide an opportunity for operators to accommodate highly scalable throughput demands in very densely populated areas through use of advanced radio technologies, one of the most promising being large Multiple-Input-Multiple-Output (MIMO) system, usually referred to as Massive MIMO (M-MIMO) [3]. M-MIMO is considered as one of enabling technologies for future cellular systems [4–6]. Studies have shown that M-MIMO is able to suppress the impacts of additive noise and uncorrelated intercell interference [3, 7]. However, gains from MIMO are highly dependent on the quality of available Channel State Information (CSI) [8]. In dense concentration of Users/User Equipment (UE), M-MIMO suffers from pilot contamination.

Optimal MIMO precoding requires CSI between each user and each Base Station (BS). A natural way to achieve this information is to use DownLink- (DL-) UpLink (UL) reciprocity of a TDD system [9–11]. The channel estimated from UL pilots can be used for precoding DL signal (and vice versa). However, M-MIMO system is characterized not only by large number of antennas but also by a large number of users. The set of orthogonal pilot sequences is usually limited, and for a large number of users we have to reuse the sequences [12]. Two users with the same pilot sequence contaminate each other’s channel estimations. Pilot contamination due to nonorthogonal training pilots has been shown to be the main capacity limiting factor of a M-MIMO system [7].

Since pilot contamination occurs due to the reuse of same pilot sequences a way to combat it is to reorthogonalize the sequences. In various papers this has been along different dimensions: such as time and space [13]. Superimposed data and pilot transmission were proposed in [14].
among BSs allows for joint processing [15] and pilot assignment [16] in order to minimize the interference.

Minimum Mean-Squared Error (MMSE) estimator is able to suppress pilot contamination from interfering channels if channel statistics of the interfering users are available. This requirement can be avoided by simply weighting pilot sequence with user specific channel coefficients that are estimated from reciprocal TDD channel [17]. The method assumes long enough coherence time for estimating the reciprocal channels in the training phase and later arranging transmission of pilots. A subspace projection using a singular value decomposition can also be used for filtering (cleaning) the interference [18, 19]. Those methods assume that receiving antennas have uncorrelated channels. Hence, the desired and the interfering signals can be projected into different subspaces based on eigenvalue decomposition of received signal matrix. The subspace based separation improves the channel estimation quality unboundedly as the number of antennas increases. For sufficiently sparse channels, simple Discrete Fourier Transform (DFT) projection can be used to remove the interference [20].

In this paper we propose data-assisted channel covariance matrix estimation algorithm for angular domain MMSE estimate in linear antenna array. In M-MIMO, such data-assisted estimation can reduce the impact of pilot contamination regardless of sparsity of the channels. Instead of statistical averages, the algorithm uses instantaneous channel power information that is extracted from the data. The algorithm works relatively well with only first data-aided steps and does not need to be iterative. However, the proposed estimator can be used as initial estimate for computationally intensive iterative algorithms such as [21].

We studied the channel estimation problem in multicell system where BSs have massive antenna arrays. The channel coherence time is allowed to be smaller than the number of BS antennas. More importantly, we assume that the BSs do not cooperate and have no explicit knowledge of channel second-order statistics. Finite-path channel model for linear antenna array [22] is considered in this work. The channel is assumed to have finite number of reflections. Angle-of-Arrivals (AoA) of multipath components are random and do not need to be orthogonal. In realistic scenario there might be very large number of dominant reflections compared to the number of receiving antennas. We introduce a codebook that projects the channel into finite (and not necessarily orthogonal) quantized beams, called angle bins. The motivation behind this approach is that the projection exposes angular sparsity of the channels. Different channels will have different power distributions over the angle bins. Channel estimates over these angle bins can be combined in such a way that pilot contamination is minimized. This is done by applying MMSE criterion to projected channel.

The paper is organized as follows. In Section 2, multicell TDD based system model is presented. Section 3 is devoted for detailed description of proposed data-aided channel estimator where practical estimation algorithms are presented. Comparison of performances of channel estimators based on numerical results is presented in Section 4. Finally, conclusions are made in Section 5.

Notation. Bold face uppercase and lowercase letters are used to denote matrices and vectors, respectively, where \( I_N \) denotes an \( N \times N \) identity matrix. Transpose, conjugate, and hermitian transpose operators are denoted by \((\cdot)^T\), \((\cdot)^\dagger\), and \((\cdot)\dagger\), respectively. \( \mathbb{E} [ \cdot ] \) denotes expectation and \( \text{tr}(\cdot) \) and \( \text{row}(\cdot) \) denote trace and row space of matrix \( \cdot \), respectively. \( \text{vec}(\cdot) \) denotes vector formed by concatenating columns of matrix \( \cdot \). \( \mathbf{h}[q] \) denotes \( q \)th element of \( \mathbf{h} \), \( \mathbf{H}[p,q] \) denotes \( p \)th row of \( \mathbf{H} \), and \( \mathbf{H}[p,q] \) denotes entry of \( \mathbf{H} \) at \( p \)th row and \( q \)th column. \( | \cdot | \) and \( \| \cdot \| \) denote absolute value and Frobenius norm, respectively. \( \hat{\cdot} \) represents definition, \( \otimes \) denotes Kronecker product, and \( \text{diag}(a_1,a_2,\ldots,a_n) \) denotes a diagonal matrix with entries \( a_1,a_2,\ldots,a_n \). Variables with bar below correspond to angular domain representations:

\[
\mathbf{X} = \mathbf{G} \mathbf{X}_{\text{est}},
\]

where \( \mathbf{G} \) is projection matrix.

2. System Model

Consider an Orthogonal Frequency Division Multiplexing (OFDM-) based multicell system with \( L \) BSs that are using the same time and frequency resources. Each BS has \( N \) antennas serving \( K \) users equipped with single antenna. All BSs are synchronized and operate in TDD fashion. Users in the same BS use orthogonal pilot codebook: \( \mathbf{S}_{P=K} = \{ \mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(K)} \} \), where \( \mathbf{s}^{(k)} \) is the pilot sequence used by the \( k \)th user. The same pilot codebook is reused in each BS. The uplink and downlink channels are reciprocal, and they are estimated from uplink pilots.

2.1. Uplink Training. While users in a cell have orthogonal pilot sequences, the same pilot sequences are reused in other cells. Received frequency-domain signal at the BS of cell 1 is given as

\[
\mathbf{Y} = \sum_{k=1}^{K} \sum_{l=1}^{L} \mathbf{h}^{(k)}_{l} \mathbf{s}^{(k)}_{l} + \mathbf{W},
\]

where \( \mathbf{h}^{(k)}_{l} \in \mathbb{C}^{N \times 1} \) is the uplink channel between the \( k \)th user in cell \( l \) and BS in cell 1 and \( \mathbf{W} \in \mathbb{C}^{N \times P} \) is complex Additive White Gaussian Noise (AWGN) having entries with zero-mean and variance \( \sigma_w^2 \). We assume that \( \| \mathbf{s}_{(k)} \|_2^2 = P \). Let us vectorize (2) as

\[
\mathbf{y} = \sum_{l=1}^{L} \sum_{k=1}^{K} \mathbf{S}^{(k)} \mathbf{h}^{(k)}_{l} + \mathbf{w},
\]

where \( \mathbf{y} = \text{vec}(\mathbf{Y}) \), \( \mathbf{S}^{(k)} = \mathbf{s}^{(k)} \otimes \mathbf{I}_N \), and \( \mathbf{w} = \text{vec}(\mathbf{W}) \).

\[
\mathbf{S}^{(k)} = \begin{pmatrix}
\mathbf{P}_{1N} & k1 = k2 \\
0 & \text{otherwise}
\end{pmatrix}.
\]

2.2. Physical Channel Model. We employ a finite-path physical channel model for linear antenna array [22]:

\[
\mathbf{h}^{(k)}_{l} = \beta^{(k)}_{l} \mathbf{A}^{(k)}_{l} \mathbf{h}^{(k)}_{l},
\]
where \( \mathbf{h}_l^{(k)} \sim \mathcal{CN}(0, \mathbf{I}_B) \) is a vector of fast-fading components from \( B \) paths and \( \rho_l^{(k)} \) is square root of the channel gain from the \( k \)-th user of cell \( l \) to the BS in cell 1 taking into account transmit power, average path-loss, and shadow fading. \( A_l^{(k)} = \{(1/\sqrt{B})a(\theta_{l1}), \cdots, (1/\sqrt{B})a(\theta_{lB})\} \in \mathbb{C}^{N \times B} \) is a matrix whose columns are \( N \times 1 \) beam vectors given by
\[
a(\theta) = \begin{bmatrix} 1, e^{-j2\pi(D/\lambda) \cos(\theta)}, \cdots, e^{-j2\pi((N-1)/D) \cos(\theta)} \end{bmatrix}^T,
\]
where \( \lambda \) is the wavelength, \( D \leq \lambda/2 \) is antenna spacing, and \( \theta \in [0, \pi] \) is AoA.

3. Angular Domain Channel Estimation

Spatial MMSE (SMMSE) estimator for MIMO systems [7, 16] requires prior knowledge of covariance matrices of the desired and interfering channels which is a difficult task. The Scaled Least-Squares (SLS) estimate [23] that needs only estimate of the Signal-to-Interference-plus-Noise Ratio (SINR) does not discriminate contaminating pilots. We propose angular domain MMSE channel estimator for antenna arrays. The angular domain channel covariance estimation is obtained by the aid of data symbols. The covariance estimation is done every Transmission Interval Time (TTI) without the need for prior information such as long-term statistics of the channel. Therefore, it is suitable for fast-fading channels.

3.1. Beam Quantization. We introduce a beam quantization codebook \( \mathbf{G} \) which is an \( N \times N \) DFT matrix. Rows of \( \mathbf{G} \) are orthogonal beams where the \( n \)-th row corresponds to an AoA \( \alpha_n = \arccos(\lambda(n-1)/ND), n = 1, \ldots, N \). Multiplying both sides of (2) with \( \mathbf{G} \) and then vectorizing like (3) give
\[
\mathbf{y} = \sum_{l=1}^{L} \mathbf{S} \mathbf{h}_l + \mathbf{w},
\]
where \( \mathbf{h}_l = \mathbf{G}_l \mathbf{y} = \text{vec}(\mathbf{G} \mathbf{Y}) \), \( \mathbf{S} = \otimes \mathbf{I}_N \), and \( \mathbf{w} = \text{vec}(\mathbf{G} \mathbf{W}) \sim \mathcal{CN}(0, \sigma_w^2 \mathbf{I}_N) \). The angular MMSE (AMMSE) estimate of \( \mathbf{h}_l \) is
\[
\mathbf{h}_{\text{AMMSE}} = \mathbf{R}_s \left( \sigma_w^2 \mathbf{I}_N + \frac{P}{B} \sum_{l=1}^{L} \mathbf{R}_s \right)^{-1} \mathbf{S}^\dagger \mathbf{y},
\]
where
\[
\mathbf{R}_s \triangleq \mathbb{E}[\mathbf{h}_l \mathbf{h}_l^T] \text{ is covariance matrix. For } N \rightarrow \infty,
\]
\[
\mathbf{h}_l[n] = \sqrt{\frac{N}{B}} \beta_l \sum_{b} \delta_{lb} \mathbf{h}_l[b],
\]
with
\[
\delta_{lb} = \begin{cases} 1, & \alpha_n = \theta_{lb} \\ 0, & \text{otherwise} \end{cases}
\]
This implies that the channel has independent entries, and hence its covariance matrix is diagonal.

For channel estimation, we approximate the covariance matrix with its M-MIMO limit (11). Therefore,
\[
\mathbf{h}_{\text{AMMSE}} = \mathbf{G}^{\dagger} \Lambda \mathbf{G} \mathbf{h}_{\text{LS}},
\]
where \( \mathbf{h}_{\text{LS}} \) is the Least-Squares (LS) channel estimate and \( \Lambda \triangleq \text{diag}(\rho_1, \ldots, \rho_N) \) is an angle bin weighting matrix which corresponds to the ratio of signal power to the total received power; we call it Fractional Signal Power (FSP). Unlike the SLS approximation of SMMSE, the AMMSE exposes the angular sparsity of the channel as illustrated in Figure 1.

3.2. Data-Aided Channel Estimation. The estimator (12) relies on average FSP. We propose an algorithm that blindly learns instantaneous FSP from the transmitted data
\[
\bar{\rho}_n = \frac{P \| \mathbf{h}_l[n] \|^2}{P \sum \| \mathbf{h}_l[n] \|^2 + (B/N) \sigma_w^2}.
\]
This done by exploiting data symbols in the estimation process. In other words, we estimate the FSP from the transmitted data symbols. The main challenge in this approach is the fact that data symbols are unknown prior to channel estimation. We solve this problem by having a two-stage channel estimation. The first stage involves a simple LS channel estimation from the strongest beam. We assume that the desired signal is stronger than interfering signals, and hence, the strongest beam is relatively less contaminated. Our algorithm uses the channel estimate from this beam to get initial estimate of data symbols. While these data symbols are likely to be erroneous, they can be used to estimate the instantaneous signal power (and hence FSP) in each beam. The accuracy of FSP estimation improves as the number of data symbols increases. Data estimation and FSP estimation make up the two stages of our algorithm. These steps can be repeated iteratively.

3.2.1. Phase 1. The goal of this phase is to have initial soft-estimate of data symbols which will be used for computing channel power. Data symbols are used due to their large number compared to pilot symbols. At initial stage, channel estimate is not available. Therefore, we rely on the strongest beam. The channel response for the strongest beam is estimated using LS. This allows us to have initial data symbols. After beam-steering, received pilot and data symbols, respectively, are
\[
\mathbf{y} = \frac{\mathbf{h}_l \mathbf{s}^T}{\text{usefull}} + \frac{\sum_{l=2}^{L} \mathbf{h}_l \mathbf{s}^T + \mathbf{w}}{\text{unwanted}},
\]
\[
\mathbf{T} = \frac{\mathbf{h}_l \mathbf{x}_l^T}{\text{usefull}} + \frac{\sum_{l=2}^{L} \mathbf{h}_l \mathbf{x}_l^T + \mathbf{w}_l}{\text{unwanted}}.
\]
where \( \mathbf{x}_l \) is vector of \( M \) data symbols transmitted from the user in cell \( l \)-th. The data symbols from the desired user, \( \mathbf{x}_1 \), will be decoded from the strongest beam using (17).
We assume that the desired channel is stronger than other interfering channels. Hence, with high probability, sum of absolute values of each row of $\mathbf{T}$ is maximized where $\mathbf{h}_1$ has largest entry. The angle bin where largest entry of $\mathbf{h}_1$ falls is found as follows. Compute $\mathbf{c}$, where

$$ c[n] = \sum_{m} |T[n,m]| $$

such that $n_{\text{max}} = \arg \max_n c[n]$ is the index of $\mathbf{h}_1$ with maximum amplitude. Channel estimate for the $n_{\text{max}}$th angle bin becomes

$$ \tilde{h}_{\text{max}} = \mathbf{h}_1[n_{\text{max}}] = \frac{1}{P} \mathbf{Y}[n_{\text{max}}] \mathbf{s}^* $$

Now we can have initial estimate of data symbols

$$ \tilde{x}_1 = \sqrt{\frac{M}{\tilde{x}_1^H \tilde{x}_1}} \tilde{x}_1, $$

where $\tilde{x}_1^H = T[n_{\text{max}} :] / \tilde{h}_{\text{max}}$ is equalized data.

3.2.2. Phase 2. We employ the initial data estimate, $\tilde{x}_1$, to find FSP. The FSP estimation is split into two parts: signal power and interference-plus-noise power estimation. The former is done by correlating the estimated data $\tilde{x}_1$ with received signal on each angle bin. Correlation of $\tilde{x}_1$ with received symbols on $n$th angle bin ($n$th row of $\mathbf{T}$) becomes

$$ \nu_n = \frac{1}{M} T[n :] \tilde{x}_1^* = \mathbf{h}_1[n] + e, $$

where

$$ e = \frac{1}{M} \sum_{l=2}^{L} \mathbf{h}_l[n] \tilde{x}_1^T \tilde{x}_1^* + \frac{1}{M} \mathbf{W}[n :] \tilde{x}_1^* $$

is the estimation error which vanishes as $M \to \infty$ such that we take $|\nu_n|$ as estimate of signal power on $n$th angle bin.
Now, consider LS channel estimate from pilots

\[
\frac{1}{P} Y[n :] s^* = h_1[n] + \sum_{l=2}^{L} h_l[n] + \frac{1}{P} W[n :] s^*. \tag{20}
\]

Subtracting (18) from (20) leaves the interfering channels and the noise such that we can estimate interference-plus-noise power as

\[
\zeta_n = \left| \frac{1}{P} Y[n :] s^* - v_n \right|^2 = \left| \sum_{l=2}^{L} h_l[n] + \frac{1}{P} W[n :] s^* - e \right|^2. \tag{21}
\]

The approximate FSP on \( n \)th angle bin is given as

\[
\hat{\rho}_n = \frac{|v_n|^2}{|v_n|^2 + \zeta_n}. \tag{22}
\]

Hence, we can construct the weighting matrix for the estimator \( \tilde{\hat{h}}_1 = \text{diag}(\hat{\rho}) \) as

\[
\hat{\Lambda} = \text{diag}(\hat{\rho}). \tag{23}
\]

4. Numerical Simulation

We evaluated performances of channel estimators for a hexagonal cellular structure with one tier of neighboring cells (see Figure 2). In order to study the impact of pilot contamination, we assume that one user in each cell located at \( 2/3 \) of the cell radius transmits the same pilot sequence. All users synchronously transmit pilot symbols followed by data symbols within the channel coherence time and coherence bandwidth. The BS in the central cell receives signals from the desired user as well as interference from users of other cells. All the interfering users are three times as far from the BS as the desired user such that \( \beta_l = 3^{-\alpha/2} \beta_1, l = 2, 3, 4, 5, 6, 7 \). Unless mentioned explicitly, simulation parameters shown in Table 1 are used.

We used normalized MSE and average uplink rate with Maximum-Ratio-Combining (MRC) as performance metrics.

4.1. MRC. MRC is a linear detector, and hence we consider only one data symbol transmission such that the received signal is expressed as

\[
t = \sum_{l=1}^{L} h_l x_l + w_t, \tag{24}
\]

where \( w_t \sim \mathcal{C}\mathcal{N}(0, \sigma_w^2 I_N) \) is AWGN and \( x_l \) is the symbol transmitted from the user in \( l \)th cell. The ergodic uplink rate for MRC receiver is given as

\[
R_1 = \mathbb{E} \left[ \log_2 (1 + \gamma_1) \right], \tag{25}
\]

where \( \gamma_1 \) is the uplink SINR.

\[
\gamma_1 = \frac{\| \hat{h}_1^\dagger h_1 \|^2}{\sum_{l=2}^{L} \| \hat{h}_l^\dagger h_l + \hat{h}_1^\dagger w_t \|^2}. \tag{26}
\]

In the Appendix we show that the uplink rate has the following upper bound:

\[
R_1 \leq \log_2 \left( 1 + \frac{2N \beta_1^2}{\sum_{l=2}^{L} \beta_l^2 + \sigma_w^2} \right) \equiv \overline{R}_1. \tag{27}
\]

4.2. Results. For a large linear antenna array, interfering channels lie on orthogonal subspace of the signaling channel
if their AoAs do not overlap [24]. In this case, MMSE estimate is interference-free. On the other hand, for overlapping AoAs, MMSE estimate is corrupted due to pilot contamination. In practical scenarios, the AoAs of desired and interfering channels can overlap. To examine both extremes, we consider two types of AoA distributions: (i) uniform: AoAs of all the users are independent and uniformly distributed over \([0, \pi]\) and (ii) directed: AoAs a user’s channel are concentrated in a narrow beam that has a width of 30° such that \(\theta_b \sim \mathcal{U}(\phi_b - \pi/12, \phi_b + \pi/12), b = 1, \ldots, B\), where \(\phi_b \sim \mathcal{U}(0, \pi)\).

When all the AoAs of all the users are narrow (directed), there is less chance of overlap between users. In Figure 3 the normalized MSE of DA-AMMSE with single and multiple iterations is illustrated. SMMSE uses channel covariance knowledge that is not available at the receiver and therefore is just a bound. DA-AMMSE performs much better than the SLS approach. The gap between DA-AMMSE and the ideal SMMSE is small. The average uplink rate shown in Figures 4 and 5 confirms this claim. The uplink rate for DA-AMMSE assuming known FSP indicates that the loss due to FSP estimation error is small. Figure 5 further reveals that when the number of reflections is large, DA-AMMSE performs almost as good as SMMSE.

Pilot contamination problem becomes worse in a rich scattering environment where covariance matrices of the desired and other interfering channels span overlapping subspaces. As can be seen from Figures 6 and 7, the uplink rate for uniform AoA distribution is far from ideal due to channel estimation error. When the number of antennas is sufficiently larger than the number of reflections, DA-AMMSE has slightly poor performance compared to the SMMSE accounting for covariance estimation error. Interestingly, when there are more reflections than the spatial dimension, DA-AMMSE outperforms the SMMSE. This indicates that the gain from instantaneous signal power based angular covariance estimation is higher than the loss due to covariance estimation error. Moreover, simulation results prove that DA-AMMSE converges in single iteration.

5. Conclusion

We studied the impact of pilot contamination in a multicell environment with noncooperative BSs having large number of antenna arrays. Then a new practical channel covariance matrix estimation algorithm for angular domain MMSE estimate is proposed. We exploit the angular sparsity of the channels in order to estimate the channel covariance matrix by the aid of data symbols. The algorithm has two
Advantages. First, no explicit knowledge of instantaneous AoAs of all beams is required. The SLS approximation of SMMSE estimator which is based on statistical average of AoA of all beams is required. The SLS approximation of SMMSE gives almost as good performance as the ideal SMMSE. In the presence of pilot contamination caused by highly overlapping AoAs, DA-AMMSE performs even better than ideal SMMSE as our proposed algorithm is based on instantaneous channel power.

\[ R_i = \mathbb{E} \left[ \log_2 (1 + \gamma_i) \right] \]
\[ \leq \log_2 (1 + \mathbb{E} [\gamma_i]). \] (A.1)

Equation (A.2) follows from Jensen inequality. The average SINR is given as

\[ \mathbb{E} [\gamma_i] = \mathbb{E} \left\{ \frac{1}{\| h_i \|^4} \right\} \]
\[ \leq \mathbb{E} \left\{ \frac{1}{\| h_i \|^4} \right\} + \mathbb{E} \left\{ \| h_i w_i \|^2 \right\} \]
\[ \leq \frac{1}{\mathbb{E} \left\{ \| h_i \|^4 \right\}} \sum_{l=2}^{L} \beta^2 \left\{ \| \tilde{h}_{l}^{\dagger} A l \| \right\} + N \sigma_w^2 \]

where

\[ \tilde{h}_1^{\dagger} A_1 \tilde{h}_1 = \frac{1}{B} \sum_{l=1}^{B} \tilde{h}_1 [l] \sum_{j}^{B} a(\theta_{ij})^{\dagger} a(\theta_{ij}) \tilde{h}_1 [j] \]
\[ = \frac{1}{B} \sum_{l=1}^{B} \left( N \| \tilde{h}_1 [l] \|^2 \right. \]
\[ + \tilde{h}_1 [l] \sum_{j \neq l}^{B} \tilde{h}_1 [j] a(\theta_{ij})^{\dagger} a(\theta_{ij}) \). \] (A.4)

After simplifications \( \mathbb{E} [\| \tilde{h}_1^{\dagger} A_1 \tilde{h}_1 \|^2] \) is given as

\[ \frac{1}{B^2} \sum_{l=1}^{B} \left( N^2 \mathbb{E} \left\{ \| \tilde{h}_1 [l] \|^2 \right\} \right. \]
\[ + \sum_{l \neq j}^{B} \mathbb{E} \left\{ \| \tilde{h}_1 [l] \|^2 \| \tilde{h}_1 [j] \|^2 \right\} \right) \]
\[ \leq \frac{1}{B^2} \left( \sum_{l=1}^{B} \mathbb{E} \left\{ \| \tilde{h}_1 [l] \|^2 \right\} \sum_{j \neq l}^{B} \mathbb{E} \left\{ \| \tilde{h}_1 [j] \|^2 \right\} \mathbb{E} \left\{ \tilde{B}_{lj} \right\} \right). \] (A.5)

We define \( \tilde{B}_{lj} = a(\theta_{ij})^{\dagger} a(\theta_{ij}) a(\theta_{ij})^{\dagger} a(\theta_{ij}) \) such that

\[ \mathbb{E} \left\{ \tilde{B}_{lj} \right\} = \text{tr} \left( \mathbb{E} \left\{ a(\theta_{ij}) a(\theta_{ij})^{\dagger} a(\theta_{ij}) a(\theta_{ij})^{\dagger} \right\} \right) \]
\[ = \text{tr} \left( \mathbb{E} \left\{ a(\theta_{ij}) a(\theta_{ij})^{\dagger} \right\} \mathbb{E} \left\{ a(\theta_{ij}) a(\theta_{ij})^{\dagger} \right\} \right) \] (A.6)
whose lower and upper bounds can be given as $N$ and $N^2$, respectively. Hence,
\[
\mathbb{E}\left[\left\|\mathbf{h}_i^\dagger \mathbf{A}_j \mathbf{A}_i \mathbf{h}_i^*\right\|^2\right] \leq \frac{1}{B'} \sum_{j=1}^{B'} \left(2N^2 + \sum_{j=2}^{B'} N^2\right)
\]
\[
+ \frac{1}{B'} \sum_{j=2}^{B'} \sum_{j=1}^{B'} N^2 = 2N^2.
\]
Hence,
\[
\mathbb{E}\left[y_1\right] \leq \frac{2N^2 \beta_1^2}{\sum_{j=2}^{B'} \beta_j^2 \mathbb{E}\left[\sum_{j=1}^{B'} (1/B') a(\theta_j) \alpha(\theta_j) a(\theta_j) a(\theta_j)\right]} + N \sigma_w^2
\]
\[
\leq \frac{2N^2 \beta_1^2}{\sum_{j=2}^{B'} \beta_j^2 \mathbb{E}\left[\sum_{j=1}^{B'} (1/B') a(\theta_j) \alpha(\theta_j) a(\theta_j) a(\theta_j)\right]} + N \sigma_w^2
\]
\[
\leq \frac{2N^2 \beta_1^2}{\sum_{j=2}^{B'} \beta_j^2 \mathbb{E}\left[\sum_{j=1}^{B'} (1/B') a(\theta_j) \alpha(\theta_j) a(\theta_j) a(\theta_j)\right]} + N \sigma_w^2
\]
\[
\leq \frac{2N^2 \beta_1^2}{\sum_{j=2}^{B'} \beta_j^2 \mathbb{E}\left[\sum_{j=1}^{B'} (1/B') a(\theta_j) \alpha(\theta_j) a(\theta_j) a(\theta_j)\right]} + N \sigma_w^2
\]
\[
\leq 2N^2 \beta_1^2 \left(\sum_{j=2}^{B'} \beta_j^2 \right) + N \sigma_w^2.
\]
Therefore,
\[
R_1 \leq \log_2 \left(1 + \frac{2N \beta_1^2}{\sum_{j=2}^{B'} \beta_j^2 + \sigma_w^2}\right) \triangleq \overline{R}_1.
\]

**Competing Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**References**


