Revisiting Sparse Channel Estimation in Massive MIMO-OFDM Systems

Zahra Shakeri*, Batoul Taki*, André L. F. de Almeida**, Mohsen Ghassemi*, and Waheed U. Bajwa*

*Dept. of Electrical and Computer Engineering, Rutgers University, USA
**Dept. of Teleinformatics Engineering , Federal University of Ceará, Brazil

***Abstract*—This paper studies training-based sparse channel estimation in massive MIMO-OFDM systems. In contrast to prior works, the focus here is on the setup in which (training) pilot tones are spread across multiple OFDM symbols. Within this setup, two training models—termed distinct block diagonal (DBD) model and repetitive block diagonal (RBD) model—are investigated. The restricted isometry property, which leads to sparse recovery guarantees, is proven for the DBD model. Further, it is established that the RBD model, through exploitation of its tensor structure, leads to computationally simpler sparse recovery algorithms. Finally, numerical experiments are provided that compare and contrast the channel estimation performance under the two models as a function of the number of pilot tones per OFDM symbol and the total number of OFDM symbols.**

I. Introduction

Employing multiple antennas in communication systems creates multiple parallel data streams and enhances system reliability [1]. Massive MIMO systems offer many advantages such as increased data throughput and link reliability that are a result of adding extra antennas to MIMO systems [2]. In such systems, coherent signal detection and low bit-error rates rely on the channel state information available at the receiver. This requires the channel to be periodically estimated at the receiver [1], [3].

The large number of transmit (Tx) and receive (Rx) antennas in massive MIMO systems gives rise to large number of channel parameters, which require considerable spectral resources to estimate them. To reduce spectral resources used for channel estimation, many works exploit the fact that wireless channels associated with a number of scattering environments tend to be highly sparse at high signal space dimension [2], [3]. In this case, training-based channel estimation techniques, which involve transmitting known data to the receiver, can exploit the literature on sparse recovery for reduction in training spectral resources [1], [3], [4].

In this work, we study sparse channel estimation of massive MIMO-OFDM channels. Most prior works on sparse channel estimation in MIMO-OFDM systems require the (training) pilot subcarriers (tones) to be interleaved with data subcarriers within one OFDM symbol [1], [3]. But practical systems tend to spread pilot tones across multiple OFDM symbols [5]. While one might anticipate that spreading training resources across frequency and time will result in the same channel estimation performance as using the same number of resources in one OFDM symbol, no prior work has formally investigated this problem to the best of our knowledge. Specifically, let N_t denote the number of OFDM symbols and let N_f be the number of OFDM pilot tones per OFDM symbol. Then the total number of training resources is $N_{tr} = N_t N_f$ and the question we want to address is: *Does the performance of sparse channel estimation in massive MIMO-OFDM systems depend on N_{tr} alone or is it also a function of N_t and N_f?*

In order to address this question, we focus on two models for training in massive MIMO-OFDM systems. In the first model, termed the *distinct block diagonal* (DBD) model, we assume independent training data are transmitted over different pilot tones. Under this model, we show that a channel with no more than S non-zero parameters can be reliably recovered from training observations as long as $N_{tr} = \Omega(S \log^2 S \log^3 p)$, where p denotes the number of channel parameters per Rx antenna. While this result suggests that estimation of (massive) MIMO-OFDM channels is largely a function of the total number of training spectral resources N_{tr}, we rush to add that this is just a sufficient condition and it comes with a few caveats that are discussed later in the paper.

The second training model discussed in this paper is termed the *repetitive block diagonal* (RBD) model, in which same training data are transmitted across different pilot tones. The motivation for this model comes from the need to reduce computational and storage complexity at the receiver in downlink settings. Consider, for instance, the setup involving 64 Tx antennas, 4 Rx antennas, and 320 delay taps per Tx-Rx pair. This results in an 81,920-dimensional channel estimation problem at the receiver, requiring large computational and storage resources. The RBD model, however, can be formulated as a Tucker decomposition of the observations [6]. In this case, we show the channel coefficient "tensor" can be recovered using the sparse tensor recovery technique referred to as Kronecker-OMP [7], which has similar performance as the classical orthogonal matching pursuit (OMP) algorithm [8], but has significantly less computational complexity and memory requirements. While we do not derive theoretical guarantees for the RBD model, we provide numerical experiments to compare its performance to that of the DBD model.

In our numerical experiments, we also study the impact of different values of N_t, N_f, and N_{tr} on the performance of both DBD and RBD models. We further investigate the use

This work is supported in part by the NSF under award CCF-1453073 and by the ARO under award W911NF-17-1-0546.

of overcomplete DFT bases, instead of the canonical bases, to model the angles of arrival (AoA) and angles of departure (AoD) in MIMO channels. Our results show that this leads to enhanced channel estimation. This suggests that data-driven bases can be learned using methods such as dictionary learning to achieve improved channel estimation performance [9].

Notational Convention: Underlined bold upper-case, bold upper-case and lower-case letters are used to denote tensors, matrices and vectors, respectively, while non-bold lower-case letters denote scalars. Also, $\|\mathbf{v}\|_2$ and $\|\mathbf{X}\|_\infty$ denote the ℓ_2 norm of $\mathbf{v}$ and max norm of $\mathbf{X}$, respectively, while $\mathbf{X}^\top$ and $\mathbf{X}^\mathsf{H}$ denote the transpose and Hermitian of $\mathbf{X}$. Moreover, $\mathbf{I}_p$ denotes the identity matrix of size $p \times p$ and $[p] \triangleq \{1, \ldots, p\}$.

II. Problem Formulation

Consider a massive MIMO-OFDM system communicating over a broadband multipath channel $\mathcal{G}$. Let N_T and N_R denote the number of Tx and Rx antennas, respectively, that are half-wavelength spaced linear arrays. Moreover, given channel bandwidth W and symbol duration T, denote $N_0 = WT$ as the temporal signal space dimension, i.e, the number of OFDM subcarriers. Assuming $W \gg 1/\tau_{\max}$, where $\tau_{\max}$ denotes the maximum delay spread of the channel, the frequency response of channel $\mathcal{G}$ can be expressed as[1]

$$\mathcal{G}(f) = \sum_{n=1}^{N_p} \beta_n \mathbf{a}_R(\theta_{R,n}) \mathbf{a}_T^\mathsf{H}(\theta_{T,n}) e^{-j2\pi\tau_n f}, \tag{1}$$

where N_p is the number of physical paths and $\mathbf{a}_R(\theta_{R,n})$ and $\mathbf{a}_T(\theta_{T,n})$ are the receive and transmit steering vectors, respectively. Here, β_n, $\theta_{R,n}$, $\theta_{T,n}$, and τ_n denote the complex path gain, AoA, AoD, and delay associated with the n-th path, respectively. The physical channel model (1) involves a large number of parameters. This motivates a virtual channel representation $\underline{\mathbf{G}}$ of $\mathcal{G}$ that can compactly and linearly model interactions between the Tx and Rx antennas. This involves a discretized approximation of $\mathcal{G}$ by sampling the angle-delay space at Nyquist rate to obtain a 3rd-order tensor $\underline{\mathbf{G}} \in \mathbb{R}^{N_R \times N_T \times N_0}$ that can be expressed via the Tucker decomposition [10] as

$$\underline{\mathbf{G}} = \underline{\mathbf{H}} \times_1 \mathbf{A}_R \times_2 \mathbf{A}_T \times_3 \mathbf{A}_F, \tag{2}$$

where $\underline{\mathbf{H}} \in \mathbb{C}^{N_R \times N_T \times L}$ denotes the virtual channel coefficient tensor with $L \triangleq \lceil W\tau_{\max} \rceil + 1$, $\mathbf{A}_R \in \mathbb{C}^{N_R \times N_R}$, $\mathbf{A}_T \in \mathbb{C}^{N_T \times N_T}$, and $\mathbf{A}_F \in \mathbb{C}^{N_0 \times L}$ are the canonical DFT bases associated with AoA, AoD, and delay spread that are used to map $\underline{\mathbf{H}}$ to $\underline{\mathbf{G}}$ [6]. In particular, each element of $\underline{\mathbf{H}}$ can be expressed in terms of the physical propagation path parameters as

$$\underline{\mathbf{H}}(i,k,l) = \sum_{n=1}^{N_p} \beta_n f_{N_R}\left(\frac{i}{N_R} - \theta_{R,n}\right) f_{N_T}^*\left(\frac{k}{N_T} - \theta_{T,n}\right) \operatorname{sinc}(l - W\tau_n), \tag{3}$$

[1]We do not consider the Doppler spread in the scope of this work.

where $f_{N_R}(\theta_R)$ and $f_{N_T}(\theta_T)$ denote the Tx and Rx smoothing kernels defined as $f_N(\theta) \triangleq \frac{1}{N}\sum_{i=1}^{N-1} e^{-j2\pi i\theta}$, and $\operatorname{sinc}(x) \triangleq \sin(\pi x)/\pi x$.

In wideband scenarios, majority of the entries of $\underline{\mathbf{H}}$ tend to be below the noise floor. Our goal is to estimate the resulting "sparse" (or approximately sparse) channel coefficient tensor $\underline{\mathbf{H}}$ using pilot training sequences transmitted over pilot subcarriers spread across N_t OFDM symbols (we assume that the channel stays constant over N_t OFDM symbols). We specifically focus on the setting in which the same set of N_f (out of N_0) subcarriers per OFDM symbol are reserved for training purposes, resulting in a total of $N_{tr} = N_t N_f$ pilot tones. Let $\mathcal{N}_f$ denote the indices of the pilot tones per OFDM symbol and $\mathbf{F} \in \mathbb{R}^{N_f \times N_0}$ denote the subcarrier selection matrix that is comprised of rows of $\mathbf{I}_{N_0}$ corresponding to $\mathcal{N}_f$. Then, each slice of training data $\underline{\mathbf{Y}} \in \mathbb{C}^{N_R \times N_t \times N_f}$ observed at the Rx antennas after N_t symbols can be expressed as

$$\underline{\mathbf{Y}}(:,:,i) = \underline{\mathbf{H}} \times_1 \mathbf{A}_R \times_2 \mathbf{X}_i \mathbf{A}_T \times_3 \mathbf{f}_i \mathbf{A}_F + \underline{\mathbf{W}}(:,:,i), \tag{4}$$

where $i \in [N_f]$, $\mathbf{f}_i$ denotes the i-th row of $\mathbf{F}$, and $\underline{\mathbf{W}} \in \mathbb{C}^{N_R \times N_t \times N_f}$ is the additive noise tensor. Here, $\mathbf{X}_i = \mathbf{X}_i^0 \mathbf{B}$ denotes the pilot sequence transmitted over the i-th subcarrier in which $\mathbf{X}_i^0 = \{\pm 1\}^{N_t \times N_t}$ is a square orthogonal matrix and $\mathbf{B} \in \mathbb{C}^{N_t \times N_T}$ is a beamforming matrix that has unit-modulus entries with random phases. This ensures the matrix $\mathbf{X}_i \mathbf{A}_T$ will have similar norm columns.

We refer to the training model described by (4) as the distinct block diagonal (DBD) model. This model can be simplified further by assuming that $\mathbf{X}_i \triangleq \mathbf{X}$ for all $i \in [N_f]$, which reduces (4) to

$$\underline{\mathbf{Y}} = \underline{\mathbf{H}} \times_1 \mathbf{A}_R \times_2 \mathbf{X}\mathbf{A}_T \times_3 \mathbf{F}\mathbf{A}_F + \underline{\mathbf{W}}. \tag{5}$$

Using properties of the Tucker decomposition [10], we can rewrite (5) as

$$\operatorname{vec}(\underline{\mathbf{Y}}) = (\mathbf{F}\mathbf{A}_F \otimes \mathbf{X}\mathbf{A}_T \otimes \mathbf{A}_R) \operatorname{vec}(\underline{\mathbf{H}}) + \operatorname{vec}(\underline{\mathbf{W}}), \tag{6}$$

where $\otimes$ denotes the matrix Kronecker product and $\operatorname{vec}(\underline{\mathbf{Y}})$ denotes the vectorized version of $\underline{\mathbf{Y}}$. We refer to this training model as the repetitive block diagonal (RBD) model.

Our focus in this paper is addressing the question: *Can we guarantee recovery of sparse $\underline{\mathbf{H}}$ under the DBD and RBD models?* We first address this question theoretically for the DBD model in the next section. Afterwards, we focus on the numerical aspects of this question for both DBD and RBD models in Section IV. In particular, due to the fact that (5) follows the Tucker decomposition, the RBD model allows recovery of $\underline{\mathbf{H}}$ using tensor recovery techniques. Specifically, Kronecker-OMP is a method introduced in [7] that does not require explicit computation of the Kronecker-structured measurement matrix $(\mathbf{F}\mathbf{A}_F \otimes \mathbf{X}\mathbf{A}_T \otimes \mathbf{A}_R)$ in (6), thus facilitating recovering of $\underline{\mathbf{H}}$ using less computation complexity and memory requirements compared to regular OMP.

III. Sparse Channel Estimation Under the DBD Model

Let us consider the linear observation model $\mathbf{y} = \mathbf{A}\mathbf{h} + \mathrm{vec}(\underline{\mathbf{W}})$ in which $\mathbf{h}$ is S-sparse (i.e., has no more than S non-zero entries). We first describe a property that is essential for recovering $\mathbf{h}$ from $\mathbf{y}$.

Proposition 1. *Let $\mathbf{A} \in \mathbb{C}^{nk\times p}$ be a matrix with unit-norm columns. To ensure reliable recovery of an S-sparse $\mathbf{h} \in \mathbb{C}^p$ from $\mathbf{y}$, $\mathbf{A}$ has to satisfy the restricted isometry property (*RIP*) of order S, i.e., $\mathbf{A} \in \mathrm{RIP}(S, \delta_S)$ with $\delta_S \in (0, 1)$ if for all S-sparse $\mathbf{h}$,*

$$(1 - \delta_S)\|\mathbf{h}\|_2^2 \leq \|\mathbf{A}\mathbf{h}\|_2^2 \leq (1 + \delta_S)\|\mathbf{h}\|_2^2. \tag{7}$$

Notice that $\mathbf{A} \in \mathrm{RIP}(S, \delta_S)$ if we have

$$\max_{\mathcal{T}\subset[p],|\mathcal{T}|\leq S} \left\|\mathbf{A}_{\mathcal{T}}^{\mathsf{H}}\mathbf{A}_{\mathcal{T}} - \mathbf{I}_{|\mathcal{T}|}\right\|_2 \leq \delta_S, \tag{8}$$

where $\mathbf{A}_{\mathcal{T}}$ denotes the matrix consisting of columns of $\mathbf{A}$ with indices $\mathcal{T}$ and $\mathbf{I}_{|\mathcal{T}|}$ denotes the identity matrix of size $|\mathcal{T}|\times|\mathcal{T}|$. Using the non-negative function $\|.\|_{\mathcal{T},S} : \mathbb{C}^{p\times p} \to [0, \infty)$ that is defined as $\|\mathbf{P}\|_{\mathcal{T},S} \triangleq \max_{\mathcal{T}\subset[p],|\mathcal{T}|\leq S} \|\mathbf{P}_{\mathcal{T}\times\mathcal{T}}\|_2$, where $\mathbf{P}_{\mathcal{T}\times\mathcal{T}}$ is a submatrix of $\mathbf{P}$ constructed by collecting entries of $\mathbf{P}$ with indices in the set $\mathcal{T} \times \mathcal{T}$, (8) can be restated as $\left\|\mathbf{A}^{\mathsf{H}}\mathbf{A} - \mathbf{I}_p\right\|_{\mathcal{T},S} \leq \delta_S$.

We now provide a theorem that shows that a special class of *structured* matrices satisfies the RIP under certain conditions. The ensuing discussion then relates this class of matrices to the observations arising within the DBD model.

Theorem 1. *Let $\mathbf{U} \in \mathbb{C}^{p\times p}$ be a unitary matrix. Define $\mathcal{X} \triangleq \{x_{i,i'}\}$, where $i \in [mk]$, $i' \in [k']$, and $k' \triangleq p/m$ is an integer factor of p, to be a generating sequence whose elements are independent realizations of Rademacher random variables taking values ± 1 with probability $1/2$. Let $\mathbf{R} \in \mathbb{R}^{mk\times p}$ be a block diagonal row-mixing matrix with $mk \leq p$, defined as*

$$\mathbf{R} \triangleq \begin{bmatrix} \mathbf{R}_1 & 0 & \dots & 0 \\ 0 & \mathbf{R}_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{R}_m \end{bmatrix}, \tag{9}$$

where

$$\mathbf{R}_i \triangleq \begin{bmatrix} x_{(i-1)k+1,1} & \dots & x_{(i-1)k+1,k'} \\ \vdots & \ddots & \vdots \\ x_{ik,1} & \dots & x_{ik,k'} \end{bmatrix}. \tag{10}$$

Next, define $\mathbf{\Phi} \triangleq \mathbf{R}\mathbf{U}$. Further, given a subset Ω of cardinality $|\Omega| = n$ chosen uniformly at random without replacement from $[m]$, define Ω' of cardinality $|\Omega'| = nk$ with elements $\Omega' = \{(i-1)k + j, i \in \Omega, j \in [k]\}$. Also, let $\mathbf{A} \in \mathbb{C}^{nk\times p}$ be the result of sampling nk rows of $\mathbf{\Phi}$ with indices in Ω' and normalizing the resulting columns by $\sqrt{m/(kn)}$. Finally, define $\mu_{\mathbf{U}} \triangleq \sqrt{p}\max_{i,j}|u_{ij}|$ as the coherence of $\mathbf{U}$. Then, for each integer p, $S > 2$, and for any $z > 1$ and any $\delta_S \in (0, 1)$, there exist positive constants c_1 and c_2 such that if $nk \geq c_1 z\mu_{\mathbf{U}}^2 S\log^2 S\log^3 p$, then $\mathbf{A}$ satisfies $\mathrm{RIP}(S, \delta_S)$ with probability higher than

$$1 - 20\max\left\{\exp\left(-c_2\delta_S^2 z\right), p^{-1}\right\}. \tag{11}$$

Similar to [11], we prove this theorem by first assuming that the block sampling variables in Ω follow Bernoulli distribution, and then translate the results for uniform distribution. To this end, let $\boldsymbol{\xi} = \{\xi_i\}_{i=1}^m$ be independent Bernoulli random variables taking value 1 with probability n/m and let $\Omega \triangleq \{i : \xi_i = 1\}$. Also, define $\boldsymbol{\eta} = \{\eta_j\}_{j=1}^{mk} = \boldsymbol{\xi} \otimes \mathbf{1}_k$ and $\Omega' = \{j : \eta_j = 1\}$. We then have the following lemmas. In all lemmas, it is assumed that $\mathbf{A}$ is a structurally-subsampled unitary matrix, as defined in Theorem 1, generated from $\mathbf{\Phi}$ according to the Bernoulli sampling model.

Lemma 1. *We have $\mathbb{E}\left[\mathbf{A}^{\mathsf{H}}\mathbf{A}\right] = \mathbf{I}_p$.*

Proof: The proof follows from steps similar to those in [12, Lemma 3.10] after some algebraic manipulations. ∎

Lemma 2. *For any integer $p > 2$ and any $r \in [2, 2\log p]$, we have*

$$\left(\mathbb{E}\left[\|\mathbf{A}\|_{\max}^r\right]\right)^{1/r} \leq \sqrt{\frac{m}{nk}}\left(\mathbb{E}\left[\|\mathbf{\Phi}\|_{\max}^r\right]\right)^{1/r} \leq \sqrt{\frac{16\mu_{\mathbf{U}}^2\log p}{nk}}.$$

Proof: The proof relies on the Khintchine inequality [13, Lemma 4.1], and follows similar steps as in [12, Lemma 3.13]. ∎

Lemma 3. *For any integer $p > 2$ and any $\varepsilon \in (0, 1)$, we have $\mathbb{E}\left[\left\|\mathbf{A}^{\mathsf{H}}\mathbf{A} - \mathbf{I}_p\right\|_{\mathcal{T},S}\right] \leq k\varepsilon$ provided*

$$nk \geq c_3\varepsilon^{-2}\mu_{\mathbf{U}}^2 S\log^2 S\log^3 p, \tag{12}$$

for some positive constant c_3.

Proof: We have

$$\mathbb{E}\left[\left\|\mathbf{A}^{\mathsf{H}}\mathbf{A} - \mathbf{I}_p\right\|_{\mathcal{T},S}\right] \overset{(a)}{\leq} \sum_{l=1}^{k}\mathbb{E}\left[\left\|\mathbf{A}_l^{\mathsf{H}}\mathbf{A}_l - \frac{1}{k}\mathbf{I}_p\right\|_{\mathcal{T},S}\right],$$

where $\mathbf{A}_l$ denotes the matrix comprised of rows of $\mathbf{A}$ with indices $\{(i-1)k + l\}_{i=1}^m$ and (a) follows from Jensen's inequality since $\|.\|_{\mathcal{T},S}$ is a norm [12]. We can show that $\mathbb{E}\left[\left\|\mathbf{A}_l^{\mathsf{H}}\mathbf{A}_l - \frac{1}{k}\mathbf{I}_p\right\|_{\mathcal{T},S}\right] \leq \varepsilon$ using similar steps as in [12, Lemma 3.14] that takes advantage of the Rudelson-Vershynin inequality [14, Lemma 3.8]. ∎

Proof of Theorem 1: A result from [15, Section 2.3] states that if subsampled matrices from a certain class satisfy RIP with probability exceeding $1 - \zeta$ for the Bernoulli sampling model, then they also satisfy RIP with probability exceeding $1 - 2\zeta$ for the uniformly-at-random sampling model. It can be shown that this result holds for the case of our block Bernoulli and uniformly-at-random sampling models as well. Hence, it is sufficient to show that $\mathbf{A}$ satisfies $\mathrm{RIP}(\delta_S, S)$ for the block Bernoulli sampling model. We next define

$$\mathbf{Y}_i \triangleq \frac{m}{nk}\xi_i\mathbf{\Phi}_i^{\mathsf{H}}\mathbf{\Phi}_i - \frac{1}{m}\mathbf{I}_p,\ \widetilde{\mathbf{Y}}_i \triangleq \frac{m}{nk}\left(\xi_i\mathbf{\Phi}_i^{\mathsf{H}}\mathbf{\Phi}_i - \xi_i'\mathbf{\Phi}_i'^{\mathsf{H}}\mathbf{\Phi}_i'\right),$$

for $i \in [m]$. Here, $\mathbf{\Phi}_i$ denotes the matrix comprised of rows of

$\mathbf{\Phi}$ with indices $\{(i-1)k+l\}_{l=1}^{k}$, ξ_i' and $\mathbf{\Phi}_i'$ are independent copies of ξ_i and $\mathbf{\Phi}_i$, and hence, $\sum_{i=1}^{m}\widetilde{\mathbf{Y}}_i$ is a symmetric version of $\sum_{i=1}^{m}\mathbf{Y}_i$. Defining $\widetilde{\mathbf{Y}} \triangleq \left\|\sum_{i=1}^{m}\widetilde{\mathbf{Y}}_i\right\|_{\mathcal{T},S}$ and $\mathbf{Y} \triangleq \|\sum_{i=1}^{m}\mathbf{Y}_i\|_{\mathcal{T},S}$, from [13], we have for all $u>0$:

$$\mathbb{E}\big[\widetilde{\mathbf{Y}}\big] \leq 2\mathbb{E}\left[\mathbf{Y}\right], \mathbb{P}\left[\mathbf{Y} > 2\mathbb{E}\left[\mathbf{Y}\right]+u\right] \leq 2\mathbb{P}\big[\widetilde{\mathbf{Y}} > u\big]. \quad (13)$$

Hence, from Lemma 3, $\mathbb{E}\big[\widetilde{\mathbf{Y}}\big] \leq 2k\varepsilon$. We can use Lemma 2 and Markov's inequality to show that with probability exceeding $1-2p^{-1}$, $\max_i \left\|\widetilde{\mathbf{Y}}_i\right\|_{\mathcal{T},S} \leq 2SB_1$, where $B_1 \triangleq \frac{16e\mu_{\mathbf{U}}^2 \log p}{n}$. Conditioned on the event $F \triangleq \left\{\max_i \left\|\widetilde{\mathbf{Y}}_i\right\|_{\mathcal{T},S} \leq 2SB_1\right\}$, using Lemma 3 and the Ledoux-Talagrand inequality [14, Lemma 3.10], if (12) is satisfied, then for any integer $r \geq q$, any $t>0$, some absolute constant $c_4>0$, and any $\varepsilon \in (0,1/k)$:

$$\mathbb{P}\big[\widetilde{\mathbf{Y}} \geq 16qk\varepsilon + 4rSB_1 + t|F\big] < \frac{c_4^r}{q^r} + 2\exp\left(\frac{-t^2}{1024qk^2\varepsilon^2}\right).$$

Next, choose $q=\lceil ec_4\rceil$, $t=32\sqrt{q}\zeta k\varepsilon$ and $r=\lceil \frac{t}{2SB_1}\rceil$ for some $\zeta>1$, and define $c_1 \triangleq \max\left\{e\sqrt{q}, c_3\right\}$. Given $\mathbb{P}(F^c) \leq 2p^{-1}$, if $nk \geq c_1\varepsilon^{-2}\mu_{\mathbf{U}}^2 S\log^2 S\log^3 p$, then $r \geq q$ and

$$\mathbb{P}\big[\widetilde{\mathbf{Y}} \geq (16q+96\sqrt{q}\zeta)k\varepsilon\big] < \exp\left(-\frac{\sqrt{q}\zeta\varepsilon kn}{3\mu_{\mathbf{U}}^2 S\log p}\right) + 2\exp(-\zeta^2) + 2p^{-1}. \quad (14)$$

We can translate this result for $\mathbf{Y}$ using (13). If (12) is satisfied, then $\mathbb{E}\left[\mathbf{Y}\right] \leq k\varepsilon$ from Lemma 3. In this case, we get

$$\mathbb{P}\left[\mathbf{Y} \geq (2+16q+96\sqrt{q}\zeta)k\varepsilon\right] < 2\exp\left(-\frac{\sqrt{q}\zeta\varepsilon kn}{3\mu_{\mathbf{U}}^2 S\log p}\right) + 4\exp(-\zeta^2) + 4p^{-1} \overset{(b)}{<} 10\max\left\{\exp\left(-c_2\delta_S^2 z\right), p^{-1}\right\},$$

where (b) follows from defining $c_5 \triangleq 2+16q+96\sqrt{q}$ (which implies $c_5\zeta k\varepsilon > (2+16q+96\sqrt{q}\zeta)k\varepsilon$), choosing $\zeta = \frac{\delta_S}{c_5 k\varepsilon}$, and denoting $c_2 \triangleq 1/c_5$ and $z \triangleq 1/(k\varepsilon)^2$. ∎

A. Discussion

Theorem 1 implies that if $nk = \Omega(\mu_{\mathbf{U}} S\log^2 S\log^3 p)$, $\mathbf{A}$ will satisfy $\mathrm{RIP}(S,\delta_S)$ with $\delta_S = \Omega(k\varepsilon)$ for an appropriately small $\varepsilon \in (0,1)$ and an S-sparse $\mathbf{h}$ is recoverable from $\mathbf{y}$ with high probability. Notice however that for large values of k, $\delta_S > 1$ and Theorem 1 will not hold. This restriction is a limitation of our proof technique.

Connecting Theorem 1 to the MIMO-OFDM observation model in (4), let $\mathbf{y}_r \in \mathbb{C}^{N_t N_f}$, $r \in [N_R]$, denote the vectorized observation received at antenna r. Also, let the indices of the N_f pilot tones be selected uniformly at random from $[N_0]$. In this case, $\mathbf{y}_r$ can be divided into N_f blocks: $\mathbf{y}_r = \begin{bmatrix}\mathbf{y}_r(1)^\top & \dots & \mathbf{y}_r(N_f)^\top\end{bmatrix}^\top$. We can write

$$\mathbf{y}_i(r) = \mathbf{X}_i(\mathbf{f}_i\mathbf{A}_F \otimes \mathbf{A}_T)\mathbf{h}_r + \mathbf{w}_i(r),\ i \in [N_f], \quad (15)$$

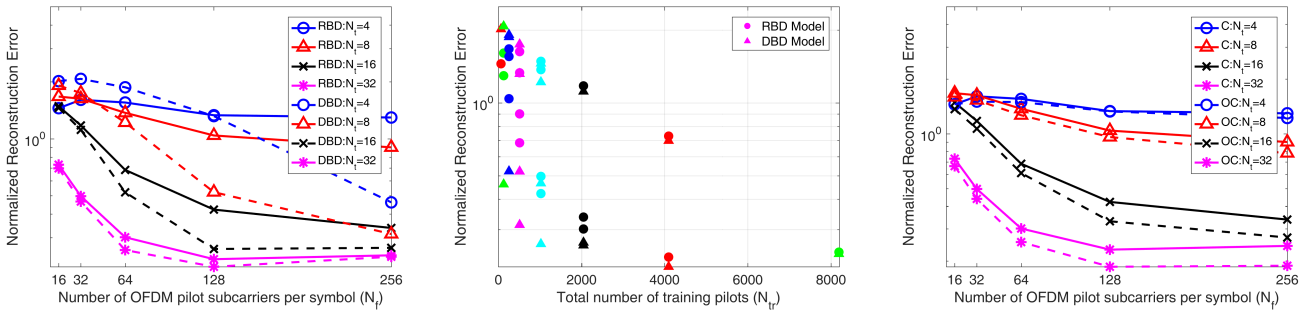
where $\mathbf{h}_r \in \mathbb{C}^{N_T L}$ is a vectorized version of channel coefficients $\underline{\mathbf{H}}(r,k,l)$, where $k \in [N_T]$ and $l \in [L]$. This corresponds to the observation model in Theorem 1 where $\mathbf{R}_i = \mathbf{X}_i$, $\mathbf{U}$ consists of stacking $\frac{1}{\sqrt{N_0}}(\mathbf{a}_{F,i} \otimes \mathbf{A}_T)$ on top of each other, where $\mathbf{a}_{F,i}$ denotes the ith row of $\mathbf{A}_F$ for $i \in [N_0]$, $k=N_t$, $n=N_f$, and $p=N_T L$. This means that for reliable recovery of $\mathbf{h}_r$, $N_t N_f = \Omega(\mu_{\mathbf{U}} S\log^2 S\log^3 N_T L)$ has to be satisfied. In comparison to the result provided in [11] for the case of $N_t=1$ that requires scaling of $N_f = \Omega(\mu_{\mathbf{U}} S\log^2 S\log^3 N_T L)$, it can be seen that the total number of parameters in $\mathbf{X}_i$, i.e. $N_{tr} = N_t N_f$, is the determining factor for reliable recovery of $\mathbf{h}_r$ in our setup. However, note that the theorem does not hold for large values of N_t since in that case $\delta_S > 1$.

Our discussion so far has been focused on sufficient conditions. In the next section, we show numerically that the sparse channel estimation performance actually depends on individual values of N_t and N_f as well as on N_{tr}.

IV. Numerical Results

In this section, we evaluate the performance of sparse channel estimation under the DBD and RBD models in terms of the number of OFDM pilot subcarriers per symbol N_f and the number of symbols N_t. The experimental setup corresponds to $N_R=4$, $N_T=64$, $N_0=1024$, $N_p=200$. We assume $W=25.12$MHz and τ_n's are uniformly distributed over $[0, 12.7\mu\text{sec}]$, resulting in $L=320$. Moreover, $(\theta_{R,n}, \theta_{T,n})$'s are uniformly distributed over $[-1/2,1/2]\times[-1/2,1/2]$ and β_n's follow the normal distribution. We use Gaussian noise with standard deviation $\sigma = 0.2\sqrt{2}$. We select the set of pilot subcarriers $\mathcal{N}_f$ uniformly-at-random from $[N_0]$. We generate random $\mathbf{X}_i$'s according to the description in Section II. We generate channel realization coefficients according to (3) and conduct experiments for $N_t = [4,8,16,32]$ and $N_f = [16,32,64,128,256]$. We use OMP and Kronecker-OMP with sparsity level $S=1000$ to reconstruct $\underline{\mathbf{H}}$ from noisy observations $\underline{\mathbf{Y}}$ for DBD and RBD models, respectively. We evaluate the channel estimation performance via the normalized reconstruction error, i.e., $\frac{\|\underline{\mathbf{G}}-\widehat{\underline{\mathbf{G}}}\|_F^2}{\|\underline{\mathbf{G}}\|_F^2}$. In all experiments, we average the error over 100 Monte Carlo experiments for random channel, additive noise, and training pilot realizations.

We conduct two sets of experiments. In both sets, the reconstruction error is plotted against N_f for various N_t's. In the first set, we compare the performance of channel estimation using the DBD and the RBD models. Figure 1(a) shows the reconstruction performance for both models (solid lines represent RBD model while dotted lines represent DBD model). For both models, it can been seen that lower error levels are achieved by increasing N_f for all N_t's. We also achieve better reconstruction when we choose a larger N_t. It can also be seen that although the DBD model outperforms the RBD model for smaller values of N_t and N_f, their performance is similar for larger values, especially for $N_f=256$. This shows that given sufficient training pilot tones, both models have a similar performance and one can use the RBD model to take advantage of Kronecker-OMP to reduce storage costs and required computational resources at the Rx. Figure 1(b) shows the error for both training models as a function of the



(a) (b) (c)

Fig. 1: Normalized reconstruction error for DBD and RBD models as a function of (a) N_f and (b) N_{tr}. In (c), we plot the normalized reconstruction error for complete (C) and overcomplete (OC) AoA and AoD bases (RBD model only).

total number of pilot tones, $N_{tr} = N_t N_f$. While it is clear that the general trend is downward based on N_{tr}, it is observed that N_{tr} is not the only determining factor and values of N_f and N_t individually matter as well in determining the error.

In the second set of experiments, we compare the performance of channel estimation using complete (C) and overcomplete (OC) bases under the RBD model. We use factor matrices $\mathbf{A}_R$, $\mathbf{X}\mathbf{A}_T$, and $\mathbf{F}\mathbf{A}_F$ to form the measurement matrix in the complete case (solid lines in Figure 1(c)) and we use overcomplete DFT matrices instead of $\mathbf{A}_R$ and $\mathbf{A}_T$ in the overcomplete setup (dotted lines in figure 1(c)). It can be observed in Figure 1(c) that the use of overcomplete DFT bases results in a reduction in the reconstruction error. This suggests that perhaps these matrices can be carefully designed using dictionary learning techniques similar to those in [9], [16] for enhanced reconstruction performance.

V. Conclusion

In this work, we studied the sparse channel estimation problem for (massive) MIMO-OFDM systems. We introduced the distinct block diagonal model for training data and obtained theoretical guarantees for channel recovery based on number of training pilot tones. Moreover, we studied the repetitive block diagonal model for training data that results in a Tucker decomposition for the observations. This formulation allows recovery of channel coefficients using sparse tensor recovery techniques that use less computational measures and memory compared to traditional recovery techniques. We further provided a comparison of the performance of the two models via numerical experiments. While our theory states that the total number of parameters determine the channel estimation performance, our numerical experiments show that the performance is also a function of the number of OFDM symbols and pilot tones. Consequently, there is a lot more that needs to be understood about the performance of these models. Future work includes providing formal guarantees for the repetitive block diagonal model using proof techniques similar to those in [17]. Our perspectives also include the use of dictionary learning techniques to improve the channel estimation performance for more challenging scenarios.

References

[1] Z. Shakeri and W. U. Bajwa, "Deterministic selection of pilot tones for compressive estimation of MIMO-OFDM channels," in *Proc. 48th Annu. Conf. Information Sciences and Systems*, 2015, pp. 1–6.

[2] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive mimo," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 104–118, 2015.

[3] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[4] W. Dongming, H. Bing, Z. Junhui, G. Xiqi, and Y. Xiaohu, "Channel estimation algorithms for broadband MIMO-OFDM sparse channel," in *14th IEEE Proc. Personal, Indoor and Mobile Radio Commun.*, vol. 2. IEEE, 2003, pp. 1929–1933.

[5] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in lte cellular networks: Key design issues and a survey," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 2, pp. 678–700, 2013.

[6] D. C. Araújo, A. L. F. de Almeida, J. P. C. L. da Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42 133–42 147, 2019.

[7] C. F. Caiafa and A. Cichocki, "Computing sparse representations of multidimensional signals using Kronecker bases," *Neural Computation*, vol. 25, no. 1, pp. 186–220, 2013.

[8] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[9] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, 2018.

[10] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," *Prob. in Measur. Change*, pp. 122–137, 1963.

[11] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "A restricted isometry property for structurally-subsampled unitary matrices," in *Proc. Annu. Allerton Conf. Commun., Control, and Computing*, 2009, pp. 1005–1012.

[12] W. U. Bajwa, "New information processing theory and methods for exploiting sparsity in wireless systems," Ph.D. dissertation, University of Wisconsin-Madison, Madison, WI, 2009.

[13] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

[14] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, 2008.

[15] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, no. 2, pp. 489–509, 2006.

[16] M. Ghassemi, Z. Shakeri, A. D. Sarwate, and W. U. Bajwa, "Learning mixtures of separable dictionaries for tensor data: Analysis and algorithms," *arXiv preprint arXiv:1903.09284*, 2019.

[17] A. Eftekhari, H. L. Yap, C. J. Rozell, and M. B. Wakin, "The restricted isometry property for random block diagonal matrices," *Appl. Comput. Harmon. Anal*, vol. 38, no. 1, pp. 1–31, 2015.