Precise Analysis of Covariance Identifiability for Activity Detection in Grant-Free Random Access

Shengsong Luo, Junjie Ma, *Member, IEEE*, Chongbin Xu, *Member, IEEE*, Xin Wang, *Fellow, IEEE*

Abstract—We consider the identifiability issue of maximum-likelihood based activity detection in massive MIMO-based grant-free random access. An intriguing observation by Chen *et al.* [1] indicates that the identifiability undergoes a phase transition for commonly-used random user signatures as L^2, N and K tend to infinity with fixed ratios, where L, N and K denote the user signature length, the total number of users, and the number of active users, respectively. In this letter, we provide a precise analytical characterization of the phase transition based on a spectral universality conjecture. Numerical results demonstrate excellent agreement between our theoretical predictions and the empirical phase transitions.

Index Terms—Random access, activity detection, phase transition, Kronecker model, spectral universality, statistical dimension.

I. INTRODUCTION

Massive machine-type communications (mMTC) involve a large number of sporadically active devices [2]–[5]. In this letter, we focus on the uplink transmission of mMTC and study the *activity detection* problem where the base station (BS), equipped with a large number of antennas, aims to identify the active devices among a bulk of potential users.

Assuming independently and identically distributed (IID) fast fading channel coefficients, the received signals at different antennas of the BS can be viewed as independent samples drawn from a common zero-mean Gaussian distribution, whose covariance matrix depends on the unknown activity pattern and the user signatures. Under such an assumption, the sample covariance matrix is a sufficient statistic for the activity pattern [6, Theorem 1]. For this reason, the maximum likelihood estimator (MLE) is also referred to as the *covariance-based approach* [6]–[8] in the literature of grant-free random access.

The large sample limit of the MLE provides important insight for its performance under massive MIMO. (Here the large sample limit corresponds to the infinite-antenna limit). It is well-known that the MLE converges (in probability) to the true parameters, provided that certain regularity conditions hold [9]. Among these conditions, the main requirement is that the model must be *identifiable*, namely, the map from the parameters to the distribution is "injective". For the non-negative MLE

S. Luo, C. Xu, and X. Wang are with the Key Lab of EMW Information (MoE), the Department of Communication Science and Engineering, Fudan University, Shanghai 200433, China (e-mails: {22110720120}@m.fudan.edu.cn, {chbinxu, xwang11}@fudan.edu.cn).

J. Ma is with the State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (e-mail: majunjie@lsec.cc.ac.cn).

Corresponding authors: Junjie Ma; Chongbin Xu.

formulation considered in [7], the identifiability of the model eventually reduces to the injectivity of a linear map under non-negativity constraint [1], which can be verified numerically using linear programming (LP) [1]. This problem has also been studied in the array signal processing literature [10], [11].

Intriguingly, Chen *et al.* [1] observed that the identifiability event undergoes a phase transition for commonly-used random user signatures, as L^2, N and K tend to infinity with fixed ratios, where L, N and K denote the signature length, the number of total users, and the number of active users, respectively. Note that the linear map involved in the identifiablity event depends on the user signatures through a complicated column-wise Kronecker product (Khatri–Rao product) structure, which poses a significant challenge for rigorous analysis of the phase transition phenomenon. In this letter, inspired by [12, Section 4.2] and the recent spectral universality principle established in [13], we introduce a surrogate model (coined "semi-random" model) for the Khatri-Rao product model, which faithfully captures the structure of the linear map and becomes mathematically tractable. Using tools developed in [14], we provide a precise analytical characterization of the phase transition phenomenon for the semi-random model. Numerical results suggest that our theoretical predictions accurately describe the empirical phase transitions for various random user signatures.

II. PROBLEM DESCRIPTION

Consider an uplink transmission from a pool of N single-antenna users to an M-antenna BS. Assume that only K of the N users are active. Each user is uniquely determined by the signature $\boldsymbol{s}_n \in \mathbb{R}^L$, $1 \leq n \leq N$. Let $u_n \in \{0, 1\}$ be an indicator variable for the activity of user n, namely, $u_n = 1$ if user n is active, and $u_n = 0$ otherwise. We model a Rayleigh fading channel experienced by the mth antenna as $h_{n,m} = \beta_n g_{n,m}$ for user n, where $g_{n,m} \sim \mathcal{N}(0, 1)$ and β_n represents the large-scale fading component. The received signal at the mth antenna is

$$\boldsymbol{y}_m = \boldsymbol{S}\text{diag}(\boldsymbol{\gamma}^{\frac{1}{2}})\boldsymbol{h}_m + \boldsymbol{w}_m, \ m = 1, \ldots, M, \quad (1)$$

where $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_N] \in \mathbb{R}^{L \times N}$ is the signature matrix, $\boldsymbol{h}_m = [h_{1,m}, \cdots, h_{N,m}]^T \in \mathbb{R}^N$ with $\boldsymbol{h}_m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $\text{diag}(\boldsymbol{\gamma}^{\frac{1}{2}})$ is a diagonal matrix whose diagonal elements are the entry-wise square root of $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_N]^T \in \mathbb{R}_+^N$ with $\gamma_n = (u_n \beta_n)^2$, and $\boldsymbol{w}_m \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

Activity detection refers to the detection of the index set of the nonzero components in $\boldsymbol{\gamma}$. Following [7], we consider a MLE framework, where $\boldsymbol{\gamma}$ is treated as a deterministic yet unknown parameter. Given $\boldsymbol{\gamma}$, the received signals $\{\boldsymbol{y}_m\}_{m=1}^M$ at all antennas are IID, following a common Gaussian distribution $\text{Pr}(\boldsymbol{y}; \boldsymbol{\gamma}) := \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \boldsymbol{S}\text{diag}(\boldsymbol{\gamma})\boldsymbol{S}^T + \sigma^2 \boldsymbol{I}$.

In this letter, we focus on the identifiability issue of the MLE. Let $\boldsymbol{\gamma}^\circ \in \mathbb{R}^N$ be the true parameter which satisfies $\|\boldsymbol{\gamma}^\circ\|_0 = K$. For computational tractability, we consider a constrained MLE formulation that takes into account the non-negativity of the parameter $\boldsymbol{\gamma}$, but not its sparsity, i.e.,

$$\max_{\boldsymbol{\gamma} \in \mathbb{R}_+^N} \prod_m \mathrm{Pr}\left(\boldsymbol{y}_m; \boldsymbol{\gamma}\right). \tag{2}$$

Since the likelihood function $\mathrm{Pr}(\boldsymbol{y}; \boldsymbol{\gamma})$ is a Gaussian distribution function with covariance $\boldsymbol{\Sigma}_\gamma = \boldsymbol{S}\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{S}^T + \sigma^2 \boldsymbol{I}$, the identifiablity (at the ground truth $\boldsymbol{\gamma}^\circ$) reduces to the following covariance identifiablity [1]:

$$\nexists \gamma \in \mathbb{R}_+^N \setminus \{\boldsymbol{\gamma}^\circ\} \implies \boldsymbol{S}\mathrm{diag}(\boldsymbol{\gamma}^\circ)\boldsymbol{S}^T = \boldsymbol{S}\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{S}^T. \tag{3}$$

This can be equivalently written as follows [1]

$$\mathcal{N}(\boldsymbol{A}) \cap \mathcal{C} = \{\boldsymbol{0}\}, \tag{4}$$

with

$$\begin{aligned}
\boldsymbol{A} &:= [\boldsymbol{s}_1 \otimes \boldsymbol{s}_1, \cdots, \boldsymbol{s}_N \otimes \boldsymbol{s}_N] \in \mathbb{R}^{L^2 \times N}, \\
\mathcal{C} &:= \{\boldsymbol{x} \in \mathbb{R}^N \mid \boldsymbol{x}_\mathcal{I} \geq \boldsymbol{0}_{N-K}\},
\end{aligned} \tag{5}$$

where $\mathcal{N}(\boldsymbol{A})$ is the null space of $\boldsymbol{A}$, $\otimes$ is the Kronecker product, and $\mathcal{I}$ is the index set for the zero entries of $\boldsymbol{\gamma}$. The cone $\mathcal{C}$ is the set of feasible directions at $\boldsymbol{\gamma}^\circ$. Let $d := \frac{L^2-L}{2}$. Specifically,

$$\mathcal{N}(\boldsymbol{A}) = \left\{\boldsymbol{x} \in \mathbb{R}^N \middle| \begin{pmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{pmatrix} \boldsymbol{x} = \boldsymbol{0} \right\}, \tag{6}$$

in which $\boldsymbol{A}_1 := \boldsymbol{S} \odot \boldsymbol{S}$, $\boldsymbol{A}_2 := [\boldsymbol{s}_1 \tilde{\otimes} \boldsymbol{s}_1, \cdots, \boldsymbol{s}_N \tilde{\otimes} \boldsymbol{s}_N] \in \mathbb{R}^{d \times N}$. Here $\odot$ is the Hadamard product, $\boldsymbol{s}_n \tilde{\otimes} \boldsymbol{s}_n (\forall n)$ denotes a d-dimensional vector which contains the lower-triangular elements of the matrix $\boldsymbol{s}_n \boldsymbol{s}_n^T$ (excluding the diagonal). We refer to [11, Definition 5] for a similar representation.

Our goal is to provide a precise phase transition characterization (in terms of the parameters L, K, N) for (4) with randomly generated $\boldsymbol{S}$.

III. SEMI-RANDOM MODEL: PHASE TRANSITION ANALYSIS

In this section, we introduce a *semi-random* model[1], which serves as a mathematically tractable model to analyze the phase transition. In Section IV, we will show that the analytical phase transition for the semi-random model accurately describes empirical results for many commonly-used signature matrices.

A. Semi-Random Model

To motivate the introduction of the semi-random model, consider a Rademacher model where the elements of $\boldsymbol{S}$ are taken from $\{\pm 1\}$. In this case, we have $\boldsymbol{A}_1 := \boldsymbol{S} \odot \boldsymbol{S} = \boldsymbol{1}_{L \times N}$, where $\boldsymbol{1}_{L \times N}$ denotes an $L \times N$ all-ones matrix; see (6). The null-space of $\boldsymbol{A}$ in (6) is the same as the null-space of $[\boldsymbol{1}_{1 \times N}; \boldsymbol{A}_2]$; where $\boldsymbol{A}_2$ is defined in (6). Motivated by this observation, and inspired by [12, Section 4.2] and the recent work of [15]–[17], we introduce the following *semi-random* matrix surrogate for $[\boldsymbol{1}_{1 \times N}; \boldsymbol{A}_2]$:

$$\boldsymbol{A}_{\text{semi-random}} := \begin{pmatrix} \boldsymbol{1}_{1 \times N} \\ \boldsymbol{A}_{\text{RI}} \end{pmatrix}, \tag{7}$$

[1]It is different from the semi-random matrix model in [15].

where $\boldsymbol{A}_{RI} \in \mathbb{R}^{d \times N}$ *is a right rotationally-invariant matrix that has the same spectrum as* $\boldsymbol{A}_2$. Here, the spectrum refers to the limiting singular value distribution of the matrix. Note that the feasibility problem in (4) is only concerned with the null space.

Universality Conjecture: In what follows, we shall focus on the semi-random model in our analysis. Our goal is to establish the precise phase transition condition for (4) using the semi-random model as a surrogate:

$$\mathcal{N}(\boldsymbol{A}_{\text{semi-random}}) \cap \mathcal{C} = \{\boldsymbol{0}\}. \tag{8}$$

B. Phase Transition Analysis

Our analysis is based on the following proposition.

Proposition 1. *[14, Theorem I]: Let* $\boldsymbol{V} \in \mathbb{R}^{N \times N}$ *be a Haar distributed random orthogonal matrix. Let* $\mathcal{D}, \mathcal{K}$ *be two closed convex cones in* $\mathbb{R}^N$. *The following holds for any* $\eta \in (0, 1)$:

$$\begin{aligned}
\frac{1}{N}\left(\delta(\mathcal{D}) + \delta(\mathcal{K})\right) \leq 1 - \xi_\eta \frac{1}{\sqrt{N}} \Rightarrow \mathrm{Pr}\{\mathcal{D} \cap \boldsymbol{V}\mathcal{K} \neq \{\boldsymbol{0}\}\} \leq \eta \\
\frac{1}{N}\left(\delta(\mathcal{D}) + \delta(\mathcal{K})\right) \geq 1 + \xi_\eta \frac{1}{\sqrt{N}} \Rightarrow \mathrm{Pr}\{\mathcal{D} \cap \boldsymbol{V}\mathcal{K} = \{\boldsymbol{0}\}\} \leq \eta,
\end{aligned} \tag{9}$$

where $\xi_\eta := \sqrt{8 \log(4/\eta)}$, *and* $\delta(\mathcal{D})$ *(similarly for* $\delta(\mathcal{K})$*) denotes the statistical dimension of* $\mathcal{D}$:

$$\delta(\mathcal{D}) := \mathbb{E}\left[\|\Pi_\mathcal{D}(\boldsymbol{g})\|^2\right], \text{ where } \boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}). \tag{10}$$

Here, $\Pi_\mathcal{D}$ *denotes the projection onto the cone* $\mathcal{D}$.

To apply Proposition 1 to analyze (8), we first rewrite (8) as

$$\mathcal{N}(\boldsymbol{A}_{\text{RI}}) \cap \mathcal{D} = \{\boldsymbol{0}\}, \tag{11}$$

where $\boldsymbol{A}_{\text{RI}}$ is a right rotationally-invariant matrix (see (7)) and

$$\mathcal{D} := \mathcal{C} \cap \{\boldsymbol{x} \mid \boldsymbol{1}^T \boldsymbol{x} = 0\} = \{\boldsymbol{x} \mid \boldsymbol{1}^T \boldsymbol{x} = 0, \boldsymbol{x}_\mathcal{I} \geq \boldsymbol{0}_{N-K}\}. \tag{12}$$

The following theorem is a direct consequence of Proposition 1, together with explicit calculations of the asymptotic of $\delta(\mathcal{D})$. Its proof can be found in the appendix.

Theorem 1. *Denote* $r := rank(\boldsymbol{A}_{RI})$. *As* $N, K, r \to \infty$ *with fixed ratios* $r/N \to \alpha \in (0, 1)$ *and* $K/N \to \epsilon \in (0, 1)$, *there exists a function* $\delta_*(\epsilon)$ *such that*

$$\begin{aligned}
\alpha > \delta_*(\epsilon) &\Rightarrow \mathrm{Pr}\{\mathcal{N}(\boldsymbol{A}_{\text{semi-random}}) \cap \mathcal{C} = \{\boldsymbol{0}\}\} \to 1, \\
\alpha < \delta_*(\epsilon) &\Rightarrow \mathrm{Pr}\{\mathcal{N}(\boldsymbol{A}_{\text{semi-random}}) \cap \mathcal{C} \neq \{\boldsymbol{0}\}\} \to 1,
\end{aligned} \tag{13}$$

where

$$\delta_*(\epsilon) := 1 - (1 - \epsilon)\Phi\left(\mu_*(\epsilon)\right), \quad \forall \epsilon \in (0, 1), \tag{14}$$

with $\mu_*(\epsilon)$ *being the unique solution to the equation of* μ *in* $(0, \infty)$:

$$(1 - \epsilon) \cdot (\mu(1 - \Phi(\mu)) - \phi(\mu)) + \epsilon\mu = 0. \tag{15}$$

Here, $\phi(\cdot)$ *and* $\Phi(\cdot)$ *denote the PDF and CDF of the standard normal distribution.*

Theorem 1 provides a precise characterization of the phase transition condition for the *identifiability* of the MLE, under a semi-random model. We see that the phase transition boundary is determined by two factors: (i) the statistical dimension of $\mathcal{D}$
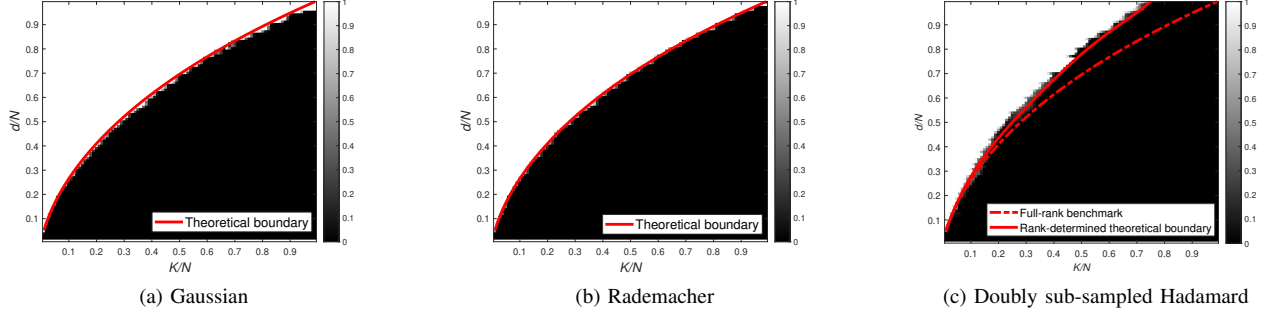
(a) Gaussian   (b) Rademacher   (c) Doubly sub-sampled Hadamard

Fig. 1: Empirical phase transition of the identifiability of MLE under various signature matrices. $d := \frac{L^2-L}{2}$ where L is length of user signatures. K is the number of active users and N is the number of total users.

(parameterized by ϵ) and (ii) the rank of $\boldsymbol{A}_{\mathrm{RI}}$. We remark that the discussions can be extended to the complex-valued cases. Details are omitted due to space limitation.

C. Behavior of the Statistical Dimension in the Sparse Limit

We analyze the behavior of the statistical dimension $\delta_*(\epsilon)$ as $\epsilon \to 0$. This provides a quantitative estimate of the sparsity-undersampling tradeoff of the MLE in the sparse limit.

Proposition 2. *Let $\delta_*(\epsilon)$ be defined as in* (14). *We have*

$$\delta_*(\epsilon) \sim 2\epsilon \log (1/\epsilon), \quad \epsilon \to 0. \tag{16}$$

Proof: Recall that $\mu_*(\epsilon) \in (0, \infty)$ is defined as the unique solution to (15), namely,

$$1 - \Phi\left(\mu_*(\epsilon)\right) - \frac{\phi\left(\mu_*(\epsilon)\right)}{\mu_*(\epsilon)} = -\frac{\epsilon}{1-\epsilon}, \quad \forall \epsilon \in (0,1). \tag{17}$$

Denote $G(\mu) := 1 - \Phi(\mu) - \frac{\phi(\mu)}{\mu}$. It is straightforward to verify that $G(\mu)$ is a monotonically increasing function on $\mu \in (0, \infty)$. Further, $\lim_{\mu \to \infty} G(\mu) = 0$ and $\lim_{\mu \to 0} G(\mu) = -\infty$. Therefore, $\mu_*(\epsilon) \to \infty$ as $\epsilon \to 0$. Towards providing a quantitative estimate of $\mu_*(\mu)$, we first note that

$$\lim_{\epsilon \to 0} \frac{1 - \Phi\left(\mu_*(\epsilon)\right)}{\epsilon \cdot \mu_*^2(\epsilon)} \overset{(a)}{=} \lim_{\epsilon \to 0} \frac{1 - \Phi\left(\mu_*(\epsilon)\right)}{-(1-\epsilon) \cdot G\left(\mu_*^2(\epsilon)\right) \cdot \mu_*^2(\epsilon)} \tag{18}$$
$$\overset{(b)}{=} 1,$$

where step (a) is from (17), step (b) is due to $\Phi(-\mu) = 1 - \Phi(\mu)$ and the following elementary fact:

$$\lim_{\mu \to \infty} \frac{1 - \Phi(\mu)}{G(\mu) \cdot \mu^2} \overset{t = -\mu}{\Longrightarrow} \lim_{t \to -\infty} \frac{\Phi(t)}{\left(\Phi(t) + \frac{\phi(t)}{t}\right) \cdot t^2} = -1.$$

Taking the logarithm of (17) and dividing both sides by $\mu_*^2(\epsilon)$ leads to ($\forall \epsilon \in (0,1)$)

$$\frac{\log \phi(\mu_*(\epsilon))}{\mu_*^2(\epsilon)} = \frac{\log\left(\mu_*(\epsilon)\right)}{\mu_*^2(\epsilon)} + \frac{\log\left(1 - \Phi\left(\mu_*(\epsilon)\right) + \frac{\epsilon}{1-\epsilon}\right)}{\mu_*^2(\epsilon)}.$$

Sending $\epsilon \to 0$ and noting that $\phi(\cdot)$ is the PDF function of the standard normal distribution, we obtain

$$\frac{\log\left(1 - \Phi\left(\mu_*(\epsilon)\right) + \frac{\epsilon}{1-\epsilon}\right)}{\mu_*^2(\epsilon)} \to -\frac{1}{2}, \quad \epsilon \to 0. \tag{19}$$

Using (18), we have

$$\frac{\log\left(1 - \Phi\left(\mu_*(\epsilon)\right) + \frac{\epsilon}{1-\epsilon}\right)}{\mu_*^2(\epsilon)} \sim \frac{\log\left(\epsilon \cdot \mu_*^2(\epsilon)\right)}{\mu_*^2(\epsilon)}, \quad \epsilon \to 0. \tag{20}$$

Combining (19) and (20) yields $\mu_*^2(\epsilon) \sim 2 \log \epsilon^{-1}(\epsilon \to 0)$. Using (18), we have $1 - \Phi(\mu_*(\epsilon)) \sim 2\epsilon \log \epsilon^{-1}$; thus[c.f. (14)]

$$\delta_*(\epsilon) := 1 - (1-\epsilon)\Phi(\mu_*(\epsilon)) \sim 2\epsilon \log \epsilon^{-1}, \quad \epsilon \to 0.$$

∎

The behavior of $\delta_*(\epsilon)$ in the sparse limit $\epsilon \to 0$ agrees with existing results on ℓ_1 minimization reconstruction for compressed sensing problems [18].

IV. SIMULATION RESULTS

This section provides numerical examples to validate the theoretical results in Theorem 1. In our experiments, we used the LP approach in [1] to certify the feasibility problem in (4). We conducted experiments under the following models of $\boldsymbol{S}$:

- *Gaussian*: the entries of $\boldsymbol{S}$ are IID standard Gaussian;
- *Rademacher*: the entries of $\boldsymbol{S}$ are IID and take $\{+1, -1\}$ with equal probability;
- *Doubly sub-sampled Hadamard*: $\boldsymbol{S}$ is a sub-sampled Hadamard matrix. Specially, the signature matrix is generated by randomly sub-sampling L rows and N columns from a $N_{\mathrm{full}} \times N_{\mathrm{full}}$ Hadamard matrix.

We set $N = 5000$ in all experiments ($N_{\mathrm{full}} = 16384$ for the doubly sub-sampled Hadamard case). The heatmap represents the empirical probability of the event (4) calculated by 100 independent realizations.

Fig. 1 displays the empirical phase transition for (4) together with its theoretical predictions in Theorem 1. The theoretical curves displayed in Fig. 1 set the values of α in the following way. Note that α is the limit of $\alpha_N := r/N$ as the problem sizes tend to infinity, where r denotes the rank of $\boldsymbol{A}_2$; see (6) for the definition of $\boldsymbol{A}_2$. For the Gaussian model and the Rademacher matrix, the matrices have full rank (i.e., $r = d$) with high probability (for the former case, with probability one). Hence, we set $\alpha = d/N$ for these two cases. For the doubly sub-sampled Hadamard model, however, the matrix $\boldsymbol{A}_2$ is rank deficient. Our numerical results suggest that $\alpha_N := r/N$ converges and we set α to this limit. Fig. 1 shows that the

theoretical results based on the semi-random model accurately describe the actual locations of the empirical phase transition for all three models. Asymptotically, the Gaussian and Rademacher signatures outperform the Hadamard signatures.

V. CONCLUSION

The phase transition analysis in this letter is based on a mathematically tractable semi-random model, which consists of an all-ones row and a rotationally-invariant random matrix. Numerical results suggest that theoretical predictions derived based on the semi-random model accurately describe the phase transition for the actual models with a complicated column-wise Kronecker product structure. Nevertheless, the validity of our analysis relies on the correctness of the universality conjecture. To this end, the recent spectral universality work [13] shed light on a possible fully rigorous proof.

APPENDIX A
PROOF OF THEOREM 1

Note that the statistical dimension of a linear subspace is exactly the subspace dimension itself. Theorem 1 is a consequence of Proposition 1 together with the following lemma.

Lemma 1. *Let $\mathcal{D}$ be defined as in (12). As $N, K \to \infty$ with $K/N \to \epsilon \in (0, 1)$, we have*

$$\frac{1}{N}\delta(\mathcal{D}) \to \delta_*(\epsilon), \tag{21}$$

where $\delta_*(\epsilon)$ is defined in (14).

Proof: We first prove that the following holds:

$$\frac{1}{N}\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^2 = \min_{\mu \in \mathbb{R}} f_N(\mu; \boldsymbol{g}), \quad \forall \boldsymbol{g} \in \mathbb{R}^N, \tag{22a}$$

where $\Pi_{\mathcal{D}}(\cdot)$ is defined in (10),

$$f_N(\mu; \boldsymbol{g}) := \frac{1}{N}\sum_{i \in \mathcal{I}}(g_i - \mu)_+^2 + \frac{1}{N}\sum_{i \in \mathcal{I}_c}(g_i - \mu)^2, \tag{22b}$$

and $\mathcal{I} := \{1, \dots, N-K\}$, $\mathcal{I}_c := \{N-K+1, \dots, N\}$. Towards proving (22), we note

$$\Pi_{\mathcal{D}}(\boldsymbol{g}) = \operatorname*{argmin}_{\boldsymbol{x}_{\mathcal{I}} \geq \boldsymbol{0}_{N-K}} \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{g}\|^2 \tag{23a}$$

$$\text{s. t.} \quad \boldsymbol{1}^T \boldsymbol{x} = 0. \tag{23b}$$

We introduce a Lagrange multiplier for the equality constraint and consider the following Lagrangian:

$$F(\boldsymbol{x}, \mu) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{g}\|^2 + \mu \boldsymbol{1}^T \boldsymbol{g}. \tag{24}$$

The minimum of $F(\boldsymbol{x}, \mu)$ over the constraint set $\{\boldsymbol{x} \in \mathbb{R}^N | \boldsymbol{x}_{\mathcal{I}} \geq \boldsymbol{0}_{N-K}\}$ admits the following closed-form expression:

$$\begin{aligned} x_i^\star &= (g_i - \mu)_+, \quad i \in \mathcal{I}, \\ x_i^\star &= (g_i - \mu), \quad i \in \mathcal{I}_c, \end{aligned} \tag{25}$$

where $(g_i - \mu)_+ = \max(g_i - \mu, 0)$. The Lagrange multiplier $\mu \in \mathbb{R}$ is chosen such that $\boldsymbol{x}^\star$ satisfies the constraint $\boldsymbol{1}^T \boldsymbol{x} = 0$:

$$\sum_{i \in \mathcal{I}}(g_i - \mu)_+ + \sum_{i \in \mathcal{I}_c}(g_i - \mu) = 0. \tag{26}$$

Note that the left-hand side (LHS) of (26) is a strictly decreasing function of μ for any $\boldsymbol{g} \in \mathbb{R}^N$. Further checking the limit behaviors of it shows that (26) has a unique solution, which we denote as μ_* in the sequel. To summarize, we have shown that $\frac{1}{N}\|\boldsymbol{x}^\star\|^2 := \frac{1}{N}\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^2$ satisfies the following

$$\begin{aligned} \frac{1}{N}\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^2 &= \frac{1}{N}\sum_{i \in \mathcal{I}}(g_i - \mu_*)_+^2 + \frac{1}{N}\sum_{i \in \mathcal{I}_c}(g_i - \mu_*)^2 \\ &= f_N(\mu_*; \boldsymbol{g}) \\ &= \min_{\mu \in \mathbb{R}} f_N(\mu; \boldsymbol{g}), \end{aligned}$$

where the last step is due to that $f_N(\mu; \boldsymbol{g})$ is a strongly convex function and its minimum is uniquely determined by its first order optimality condition, which is given by (26). This completes the proof of (22).

The characterization in (22) is a deterministic result and holds for all $\boldsymbol{g} \in \mathbb{R}^N$. We now consider $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ and prove the almost sure convergence of $\frac{1}{N}\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^2$ as $N, K \to \infty$ with $K/N \to \epsilon \in (0, 1)$. By the strong law of large numbers, the following holds for any fixed $\mu \in \mathbb{R}$:

$$f_N(\mu; \boldsymbol{g}) \xrightarrow{a.s.} f(\mu), \tag{27}$$

where

$$f(\mu) := (1 - \epsilon)\mathbb{E}\left[(G - \mu)_+^2\right] + \epsilon\mathbb{E}\left[(G - \mu)^2\right], \tag{28}$$

and $G \sim \mathcal{N}(0, 1)$. It is easy to see that $f(\mu)$ is a strongly convex function of μ and hence admits a unique minimum. By the convexity lemma [19, Theorem 10.8], pointwise convergence of a sequence of convex functions implies uniform convergence over any compact set; see also [20]. The function $f(\mu)$ has a unique minimum. Using a standard "artificial boundedness" argument as in [21, Lemma 10], we then conclude that

$$\min_{\mu \in \mathbb{R}} f_N(\mu; \boldsymbol{g}) \xrightarrow{a.s.} \min_{\mu \in \mathbb{R}} f(\mu). \tag{29}$$

By (22) and the definition of $f(\cdot)$ in (28), we have

$$\frac{1}{N}\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^2 \xrightarrow{a.s.} \min_{\mu \in \mathbb{R}} (1 - \epsilon)\mathbb{E}\left[(G - \mu)_+^2\right] + \epsilon\mathbb{E}\left[(G - \mu)^2\right]. \tag{30}$$

Recall that $\frac{1}{N}\delta(\mathcal{D}) := \mathbb{E}\left[\frac{1}{N}\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^2\right]$. To translate (30) into the convergence of $\delta(\mathcal{D})/N$, we next prove $\sup_N \mathbb{E}\left[\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^4/N^2\right] < \infty$, which guarantees the uniform integrability of $\frac{1}{N}\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^2$. By the polar decomposition [14], $\|\boldsymbol{g}\|^2 = \|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^2 + \|\Pi_{\mathcal{D}^o}(\boldsymbol{g})\|^2$, where $\mathcal{D}^o$ denotes the polar cone of $\mathcal{D}$. Hence, $\frac{1}{N}\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^2 \leq \frac{1}{N}\|\boldsymbol{g}\|^2$. When $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$, we have $\mathbb{E}[\|\boldsymbol{g}\|^4] = N^2 + N$. Hence, $\mathbb{E}\left[\|\Pi_{\mathcal{D}}(\boldsymbol{g})\|^4/N^2\right] \leq (N^2 + N)/N^2 \leq 2, \forall N \geq 1$. This verifies the uniform integrability condition; hence

$$\frac{1}{N}\delta(\mathcal{D}) \to \min_{\mu \in \mathbb{R}} (1 - \epsilon)\mathbb{E}\left[(G - \mu)_+^2\right] + \epsilon\mathbb{E}\left[(G - \mu)^2\right]. \tag{31}$$

It remains to check that the solution to the right-hand side (RHS) of (31) is the unique solution to (15). This follows from checking the first order optimality condition for (15) $-(1 - \epsilon)\mathbb{E}\left[(G - \mu)_+\right] + \mu\epsilon = 0$, which recovers (15) together with $\mathbb{E}\left[(G - \mu)_+\right] = \phi(\mu) - \mu(1 - \Phi(\mu))$, $\forall \mu \in \mathbb{R}$. It is straightforward to show that the unique solution is positive. This completes the proof. ∎

REFERENCES

[1] Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Phase transition analysis for covariance-based massive random access with massive MIMO," *IEEE Transactions on Information Theory*, vol. 68, no. 3, pp. 1696–1715, 2022.

[2] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 59–65, 2016.

[3] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the Internet of Things," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 88–99, 2018.

[4] N. H. Mahmood, S. Böcker, A. Munari, F. Clazzer, I. Moerman, K. Mikhaylov, O. Lopez, O.-S. Park, E. Mercier, H. Bartz *et al.*, "White paper on critical and massive machine type communication towards 6G," *arXiv preprint arXiv:2004.14146*, 2020.

[5] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Select. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[6] S. Haghighatshoar, P. Jung, and G. Caire, "Improved-scaling law for activity detection in massive MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2018, pp. 381–385.

[7] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2925–2951, 2021.

[8] S. Khanna and C. R. Murthy, "On the support recovery of jointly sparse gaussian sources via sparse Bayesian learning," *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7361–7378, 2022.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[10] P. Pal and P. P. Vaidyanathan, "Correlation-aware sparse support recovery: Gaussian sources," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2013, pp. 5880–5884.

[11] A. Koochakzadeh, H. Qiao, and P. Pal, "On fundamental limits of joint sparse support recovery using certain correlation priors," *IEEE Transactions on Signal Processing*, vol. 66, no. 17, pp. 4612–4625, 2018.

[12] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete & Computational Geometry*, vol. 43, pp. 522–541, 2010.

[13] R. Dudeja, S. Sen, and Y. M. Lu, "Spectral universality of regularized linear regression with nearly deterministic sensing matrices," *IEEE Transactions on Information Theory, to appear*, 2024. [Online]. Available: https://arxiv.org/abs/2208.02753

[14] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference: A Journal of the IMA*, vol. 3, no. 3, pp. 224–294, 2014.

[15] R. Dudeja, Y. M. Lu, and S. Sen, "Universality of approximate message passing with semirandom matrices," *The Annals of Probability*, vol. 51, no. 5, pp. 1616–1683, 2023.

[16] A. Abbara, A. Baker, F. Krzakala, and L. Zdeborová, "On the universality of noiseless linear estimation with respect to the measurement matrix," *Journal of Physics A: Mathematical and Theoretical*, vol. 53, no. 16, p. 164001, 2020.

[17] R. Dudeja and M. Bakhshizadeh, "Universality of linearized message passing for phase retrieval with structured sensing matrices," *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7545–7574, 2022.

[18] D. L. Donoho and J. Tanner, "Precise undersampling theorems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, 2010.

[19] R. T. Rockafellar, "Convex analysis:(pms-28)," 2015.

[20] D. Pollard, "Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, vol. 7, no. 2, pp. 186–199, 1991.

[21] C. Thrampoulidis, E. Abbasi, and B. Hassibi, "Precise error analysis of regularized m-estimators in high dimensions," *IEEE Transactions on Information Theory*, vol. 64, no. 8, pp. 5592–5628, 2018.