Phase Transition Analysis for Covariance Based Massive Random Access with Massive MIMO

Zhilin Chen and Wei Yu
Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada
Emails: {zchen, weiyu}@ece.utoronto.ca

***Abstract*—This paper studies the massive random access problem in which a large number of sporadically active devices wish to communicate to a base-station (BS) equipped with a large number of antennas. The devices are pre-assigned unique pilot sequences for random access. It has been shown previously that the device activity detection problem at the BS can be formulated as a maximum likelihood estimation (MLE) problem, whose solution depends on the sample covariance matrix of the received signal. This paper adopts the MLE formulation, and proposes an approach to analyze the covariance based detection by studying the asymptotic properties of the MLE via its associated Fisher information matrix. This paper proposes a necessary condition on the Fisher information matrix such that the estimation error tends to zero in the massive multiple-input multiple-output (MIMO) regime. A phase transition analysis is carried out based on the necessary condition. This paper also analyzes the distribution of the estimation error for the case with a large but finite number of antennas at the BS. Numerical experiments validate the analysis.**

I. Introduction

Random access is a challenging task for massive machine-type communications (mMTC), where a large number of sporadically active devices communicate with the base-station (BS) in the uplink. Conventional grant-based random access with orthogonal sequences is not suitable for mMTC due to the fact that the number of potential devices often greatly exceeds the number of time-frequency dimensions. Instead, a grant-free based random access strategy is envisioned to be a promising approach [1], [2], in which each active device directly transmits the pilot and the data to the BS without waiting for the grant.

There are various designs for the grant-free based random access. Most of them rely on the use of non-orthogonal sequences to accommodate massive number of devices. The non-orthogonal sequences can be used as signatures in active device identification, e.g., [3]–[5], as codewords in information transmission, e.g., [6]–[8], or as both [2], [9]. The common idea is that in the random access phase, each active device transmits a non-orthogonal sequence to the BS, and by detecting which sequences are transmitted, the BS acquires the identification of the active devices, or/and the information bits.

Non-orthogonal sequence detection at the BS is a crucial task in mMTC. The sequence detection problem is closely related to a sparse signal recovery problem due to the sporadic nature of the device activity. In particular, the sequence detection problem can be formulated as a compressed sensing problem, for which a variety of techniques can be explored. For example, the use of the efficient approximate message passing (AMP) has been proposed for the massive device activity detection problem in [3], [4]. An important feature of AMP is that there is an analytical framework called state evolution [10] for performance analysis, based on which the probabilities of false alarm and missed detection can be accurately predicted.

As an alternative to AMP, for the massive multiple-input multiple-output (MIMO) systems, a covariance based approach for device activity detection is proposed in [5], in which the sequence detection problem is formulated as either a maximum likelihood estimation (MLE) problem, or a non-negative least square (NNLS) problem. In both cases, the optimization problem is formulated in terms of certain covariance matrix. The covariance based method is used in [5] for device activity detection and in [8] for data decoding, where it is shown that the covariance based method with the MLE formulation outperforms AMP in the massive MIMO regime. In [11], a reconstruction error bound and a scaling law on the system parameters for the NNLS formulation are provided. It is also shown in [11] that the same scaling law applies to a binary constrained version of the MLE formulation.

In this paper, we consider the device activity detection problem in the massive MIMO setup similar to [5] and [11], and analyze the performance of the covariance based method under the general MLE formulation. We provide a numerical phase transition analysis by exploiting the asymptotic properties of the MLE with non-negative constraint, and by studying its associated Fisher information matrix.

As compared to our previous work [9] that deals with only non-singular Fisher information matrix, this paper considers general Fisher information matrix, based on which a new necessary condition to make the estimation error to approach zero in the massive MIMO regime is proposed. The condition, which involves solving a linear programming (LP) problem, helps identify a phase transition in the system parameter space to differentiate the success region and the failure region for the covariance based approach. As compared to the scaling law in [11] that analyzes the asymptotic order of the number of identifiable active devices as a function of the pilot length, the phase transition analysis in this paper is based on solving an LP, so it is numerical in nature; but it is non-asymptotic in the number of active devices. Moreover, this paper characterizes the distribution of the estimation error using the general Fisher information matrix. The characterization, which involves solving a quadratic programming (QP) problem, accurately predicts the error probabilities in device activity detection.

II. System Model

Consider an uplink single-cell massive MIMO system with M antennas at the BS and N single-antenna devices. We assume that only $K \ll N$ devices are active during a time slot. For the purpose of active device identification, suppose that each device n in the system maintains a unique signature sequence $\mathbf{s}_n = [s_1, \ldots, s_L]^T \in \mathbb{C}^{L\times 1}$, where L is the length of the sequence and is assumed to be shorter than the channel coherence length. When device n is active during a time slot, it transmits sequence $\mathbf{s}_n$ to the BS as a random access request.

Let $a_n \in \{1, 0\}$ denote the activity of device n in a given time slot. Let $g_n\mathbf{h}_n$ denote the channel vector between the BS and device n, where $\mathbf{h}_n \in \mathbb{C}^{M\times 1}$ is the Rayleigh fading component following independent and identically distributed (i.i.d.) complex Gaussian distribution with zero mean and unit variance, and g_n is the large-scale fading component including path-loss and shadowing. We assume block fading channel, i.e., the channels are constant during a coherence block, and assume that the sequences are transmitted synchronously. The received signal $\mathbf{Y} \in \mathbb{C}^{L\times M}$ at the BS can be expressed as

$$\mathbf{Y} = \sum_{n=1}^{N} a_n \mathbf{s}_n g_n \mathbf{h}_n^T + \mathbf{W} \triangleq \mathbf{S}\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H} + \mathbf{W}, \tag{1}$$

where $\mathbf{S} \triangleq [\mathbf{s}_1, \ldots, \mathbf{s}_N] \in \mathbb{C}^{L\times N}$ is the sequence matrix, $\boldsymbol{\Gamma} \triangleq \mathrm{diag}\{\gamma_1, \ldots, \gamma_N\} \in \mathbb{R}^{N\times N}$ with $\gamma_n = a_n g_n^2$ is a diagonal matrix that indicates both the device activity and the large-scale fading components, $\mathbf{H} \triangleq [\mathbf{h}_1, \ldots, \mathbf{h}_N]^T \in \mathbb{C}^{N\times M}$ is the channel matrix, and $\mathbf{W} \in \mathbb{C}^{L\times M}$ is the effective i.i.d. Gaussian background noise whose variance σ_w^2 is normalized by the device transmit power for simplicity. We use $\boldsymbol{\gamma} \triangleq [\gamma_1, \ldots, \gamma_N]^T \in \mathbb{R}^{N\times 1}$ to denote the diagonal entries of $\boldsymbol{\Gamma}$.

We assume that all signature sequences are known at the BS, and all sequences are generated from i.i.d complex Gaussian distribution with zero mean and unit variance.

III. Covariance Based Device Activity Detection

A. Problem Formulation

The BS identifies the active devices by detecting a_n from the received signal $\mathbf{Y}$. Since a_n is contained in the diagonal entries of $\boldsymbol{\Gamma}$ in (1), the detection of a_n can be formulated as the estimation of $\boldsymbol{\Gamma}$ (or $\boldsymbol{\gamma}$), by exploiting the statistics of the channels and the background noise. Once an estimate $\hat{\boldsymbol{\gamma}}$ is obtained, the binary indicator a_n can be determined through simple operations on $\hat{\boldsymbol{\gamma}}$, e.g., element-wise thresholding.

It is worth noting that if channel estimation is needed in addition to device activity detection, we can use an alternative compressed sensing approach treating $\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}$ as a row sparse matrix. However, such a problem involves many more parameters and is therefore more challenging to solve.

Following the approach suggested in [5], we estimate $\boldsymbol{\gamma}$ from $\mathbf{Y}$ using the MLE. To compute the likelihood, we first observe from (1) that given $\boldsymbol{\gamma}$, the columns of $\mathbf{Y}$, denoted by $\mathbf{y}_m \in \mathbb{C}^{L\times 1}, 1 \le m \le M$, are independent due to the i.i.d. channels, and each column follows a multivariate complex Gaussian distribution as

$$\mathbf{y}_m \sim CN\left(\mathbf{0}, \mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma_w^2\mathbf{I}\right), \tag{2}$$

where the mean and the covariance are obtained by computing $\mathbb{E}[\mathbf{y}_m\mathbf{y}_m^H]$ from (1). Let $\boldsymbol{\Sigma} \triangleq \mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma_w^2\mathbf{I}$. Due to the independence of the columns of $\mathbf{Y}$, the likelihood of $\mathbf{Y}$ is

$$\begin{aligned} p(\mathbf{Y}|\boldsymbol{\gamma}) &= \prod_{m=1}^{M} \frac{1}{|\pi\boldsymbol{\Sigma}|} \exp\left(-\mathbf{y}_m^H\boldsymbol{\Sigma}^{-1}\mathbf{y}_m\right) \\ &= \frac{1}{|\pi\boldsymbol{\Sigma}|^M} \exp\left(-\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{Y}\mathbf{Y}^H\right)\right), \end{aligned} \tag{3}$$

where $|\cdot|$ and $\mathrm{tr}(\cdot)$ denote the determinant and the trace of a matrix, respectively. The maximization of $\log p(\mathbf{Y}|\boldsymbol{\gamma})$ can be cast as the minimization of $-\frac{1}{M}\log p(\mathbf{Y}|\boldsymbol{\gamma})$ expressed as

$$\underset{\boldsymbol{\gamma}}{\text{minimize}} \quad \log|\boldsymbol{\Sigma}| + \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}\right) \tag{4a}$$

$$\text{subject to} \quad \boldsymbol{\gamma} \ge 0, \tag{4b}$$

where $\hat{\boldsymbol{\Sigma}} \triangleq \frac{1}{M}\mathbf{Y}\mathbf{Y}^H$ is the sample covariance matrix of the received signal averaged over different antennas, and the constraint $\boldsymbol{\gamma} \ge 0$ is due to the fact $\gamma_n = a_n g_n^2 \ge 0$, which defines a natural parameter space of $\boldsymbol{\gamma}$.

We observe from (4) that the MLE depends on $\mathbf{Y}$ through the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$. As M increases, $\hat{\boldsymbol{\Sigma}}$ will tend to the true covariance matrix $\boldsymbol{\Sigma}$, but the size of the problem in (4) does not change, which makes it preferred in the massive MIMO regime. Meanwhile, due to the averaging operation over antennas in the sample covariance, the channel hardening effect offered by massive MIMO is exploited.

B. Algorithms

The optimization problem (4) is not convex in general due to the fact that $\log|\boldsymbol{\Sigma}|$ is concave whereas $\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}})$ is convex. However, various algorithms have shown excellent performance in practice for solving (4). For example, the authors of [12] propose a multiple sparse Bayesian learning (M-SBL) algorithm based on expectation maximization that estimates $\boldsymbol{\gamma}$ iteratively. The authors of [5] suggest a coordinate descent algorithm that randomly updates each coordinate of the estimate $\hat{\boldsymbol{\gamma}}$ until convergence. Although the problem is non-convex, the global optimality of the solution by M-SBL or coordinate descent for such a problem can be justified if $\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{H}$ or $\mathbf{S}$ satisfies certain conditions; see [12] and [5].

In this paper, we adopt the coordinate descent method from [5] to solve (4) in the simulations. Once an estimate $\hat{\boldsymbol{\gamma}}$ is obtained, we can employ the element-wise thresholding to determine a_n from $\hat{\gamma}_i$ with some common threshold l_{th}. The probabilities of missed detection and the false alarm can be traded off by setting different values of the threshold.

IV. Asymptotic Performance Analysis

The goal of this paper is to analyze the performance of the covariance based device activity detection. The analysis is based on the characterization of the solution $\hat{\boldsymbol{\gamma}}$ to problem

(4) in the regime $M \to \infty$. We aim to investigate two questions: (i) What are the conditions on the system parameters L, N, K, σ_w^2 such that the estimate $\hat{\gamma}$ can approach the true parameter γ^0 as $M \to \infty$? (ii) How is the estimation error $\hat{\gamma} - \gamma^0$ distributed, if those conditions are satisfied but under finite M? The first question helps identify the desired operating regime in the space of L, N, K, σ_w^2 for getting an accurate estimate $\hat{\gamma}$ via the MLE with massive MIMO, and the second one helps characterize the error probabilities.

We investigate these two questions by exploiting the asymptotic properties of the MLE: *consistency* and *asymptotic normality*. Recall from [13] that under certain regularity conditions, as the number of i.i.d. samples increases, the estimate $\hat{\gamma}$ is *consistent*, i.e.,

$$\hat{\gamma} \xrightarrow{P} \gamma^0, \quad \text{as } M \to \infty, \tag{5}$$

where $\xrightarrow{P}$ denotes convergence in probability. Furthermore, if the true parameter γ^0 is an interior point in the parameter space of γ, the estimation error $M^{\frac{1}{2}}(\hat{\gamma} - \gamma^0)$ tends to a Gaussian distribution as the number of i.i.d. samples increases, i.e.,

$$M^{\frac{1}{2}}(\hat{\gamma} - \gamma^0) \xrightarrow{D} \mathcal{N}(0, M\mathbf{J}^{-1}(\gamma^0)), \quad \text{as } M \to \infty, \tag{6}$$

where $\xrightarrow{D}$ denotes convergence in distribution, and $\mathbf{J}(\gamma)$ is the Fisher information matrix, whose (i,j)-th entry is defined as

$$[\mathbf{J}(\gamma)]_{ij} = \mathbb{E}\left[\left(\frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_i}\right)\left(\frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_j}\right)\right], \tag{7}$$

where $p(\mathbf{Y}|\gamma)$ is given in (3), and the expectation is taken with respect to $\mathbf{Y}$.

However, for the MLE considered in this paper, the results in (5) and (6) cannot be directly applied as two of the regularity conditions may be violated: (i) the consistency of the MLE requires that the true parameter γ^0 is *identifiable*, i.e., there exists no other $\gamma\prime \neq \gamma^0$ such that $p(\mathbf{Y}|\gamma') = p(\mathbf{Y}|\gamma^0)$. This may not be the case in our considered problem because the dimension of γ^0 could be very large and ambiguity may occur; (ii) the asymptotic normality of the MLE requires that the true parameter γ^0 is an interior point in its parameter space, which is $[0, +\infty)^N$, but in our problem γ^0 in fact lies on the boundary as most of the entries in γ^0 are zero. The boundary condition makes the estimation error $\hat{\gamma} - \gamma^0$ always non-negative at some coordinates, instead of being Gaussian.

In this paper, we deal with the consistency issue by proposing a new necessary condition for the parameter identifiability, and deal with the asymptotic distribution of $M^{\frac{1}{2}}(\hat{\gamma} - \gamma^0)$ by taking the boundary case into consideration. Since $\mathbf{J}(\gamma)$ plays a key role in our analysis, we first provide an explicit expression of $\mathbf{J}(\gamma)$, which has been derived in [9], as follows.

Proposition 1. *Consider the likelihood function in* (3)*, where γ is the parameter to be estimated. The Fisher information matrix for estimating γ is*

$$\mathbf{J}(\gamma) = M\left(\mathbf{P} \odot \mathbf{P}^*\right), \tag{8}$$

where $\mathbf{P} \triangleq \mathbf{S}^H\left(\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{S}$, $\odot$ is the element-wise product, and $(\cdot)^*$ is the conjugate operation.

It is worth noting that $\mathbf{J}(\gamma)$ in (8) may be singular depending on the values of L and N. This can be seen by checking the rank of $\mathbf{J}(\gamma)$. Using $\mathrm{Rank}(\mathbf{U} \odot \mathbf{V}) \leq \mathrm{Rank}(\mathbf{U})\,\mathrm{Rank}(\mathbf{V})$ for arbitrary matrices $\mathbf{U}$ and $\mathbf{V}$ and $\mathrm{Rank}(\mathbf{P}) \leq L$, we get

$$\mathrm{Rank}(\mathbf{P} \odot \mathbf{P}^*) \leq \mathrm{Rank}(\mathbf{P})^2 \leq L^2. \tag{9}$$

Since $\mathbf{P} \odot \mathbf{P}^*$ is of size $N \times N$, $\mathbf{J}(\gamma)$ is singular if $N > L^2$, i.e., the dimension of γ is larger than the size of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ in (4). The singularity of $\mathbf{J}(\gamma)$ complicates the analysis of the estimation problem. Our analysis below takes the singular $\mathbf{J}(\gamma)$ into consideration.

A. A Necessary Condition for Consistency of $\hat{\gamma}$

We first establish a necessary condition on $\mathbf{J}(\gamma)$ such that $\hat{\gamma}$ can approach γ^0 in the large M limit.

Theorem 1. *Let γ^0 denote the true parameter, and let $\hat{\gamma}$ denote the solution of* (4)*. Let $\mathcal{I}$ be an index set corresponding to the zero entries of γ^0, i.e., $\mathcal{I} \triangleq \{i \mid \gamma_i^0 = 0\}$. We define two sets $\mathcal{N}, \mathcal{C}$ in the space $\mathbb{R}^N$, respectively, as follows*

$$\mathcal{N} \triangleq \{\mathbf{x} \mid \mathbf{x}^T\mathbf{J}(\gamma^0)\mathbf{x} = 0, \mathbf{x} \in \mathbb{R}^N\},$$
$$\mathcal{C} \triangleq \{\mathbf{x} \mid x_i \geq 0, i \in \mathcal{I}, \mathbf{x} \in \mathbb{R}^N\},$$

where x_i is the i-th entry of $\mathbf{x}$. Then a necessary condition for the consistency of $\hat{\gamma}$, i.e., $\hat{\gamma} \to \gamma^0$ as $M \to \infty$, is that the intersection of $\mathcal{N}$ and $\mathcal{C}$ is zero, i.e., $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$.

We define $\mathcal{N}, \mathcal{C}$ in $\mathbb{R}^N$ since $\mathbf{J}(\gamma)$ is real. An interpretation of the condition $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ is as follows. Set $\mathcal{N}$ is a subspace in $\mathbb{R}^N$ spanned by the eigenvectors of $\mathbf{J}(\gamma^0)$ corresponding to zero eigenvalues, i.e., the null space of $\mathbf{J}(\gamma^0)$. Set $\mathcal{C}$ is a cone with the coordinates indexed by $\mathcal{I}$ being non-negative. Base on $\mathcal{C}$, we can express the neighborhood of γ^0 in the parameter space $[0, +\infty)^N$ as $\gamma^0 + t\mathbf{x}$ for any $\mathbf{x} \in \mathcal{C}$ and some positive scalar t. The condition says that any direction from γ^0 to the feasible neighborhood cannot lie in the null space of $\mathbf{J}(\gamma^0)$.

Proof: We prove this by contradiction. We show that if there exists a non-zero vector $\mathbf{x} \in \mathcal{N} \cap \mathcal{C}$, then the likelihood function $p(\mathbf{Y}|\gamma)$ stays unchanged when γ moves from γ^0 to the neighborhood along the direction $\mathbf{x}$. The unchanged $p(\mathbf{Y}|\gamma)$ indicates that the true parameter γ^0 cannot be uniquely identified around its neighborhood based on $p(\mathbf{Y}|\gamma)$. Therefore, it cannot be guaranteed that the estimate $\hat{\gamma}$ obtained from MLE arbitrarily approaches the true parameter γ^0.

To prove the result, we note that $\mathbf{x}$ satisfies $\mathbf{x}^T\mathbf{J}(\gamma^0)\mathbf{x} = 0$ since $\mathbf{x} \in \mathcal{N}$. By plugging (7) into $\mathbf{x}^T\mathbf{J}(\gamma^0)\mathbf{x}$, we get

$$\mathbf{x}^T\mathbf{J}(\gamma^0)\mathbf{x} = \mathbb{E}\left[\sum_i \frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_i} x_i\right]^2_{\gamma=\gamma^0} = 0. \tag{10}$$

By noting that the term in the middle is non-negative, we get

$$\sum_i \frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_i} x_i\Big|_{\gamma=\gamma^0} = 0, \tag{11}$$

from which we conclude that $\log p(\mathbf{Y}|\gamma)$ stays unchanged when γ moves from γ^0 to its neighborhood along the direction $\mathbf{x}$, which implies non-identifiability. ■

Note that to establish Theorem 1, we make use of the notion that the true parameter $\boldsymbol{\gamma}^0$ should be uniquely identified by the likelihood function in its neighborhood. Such a property is often referred to as the local identifiability [14]. The condition is necessary because $\boldsymbol{\gamma}^0$ can only be locally identified. To prove sufficiency, one would need to establish that it is identifiable *globally* in the whole parameter space.

In the special case when $\mathbf{J}(\boldsymbol{\gamma}^0)$ is non-singular, which is often true if $N \leq L^2$ from (8), we have $\mathcal{N} = \{\mathbf{0}\}$, and the condition in Theorem 1 is immediately satisfied.

Since there is no closed-form expression of $\mathcal{N} \cap \mathcal{C}$ in general, the condition $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ for a given $\mathbf{J}(\boldsymbol{\gamma}^0)$ cannot be verified analytically. However, by noting that both $\mathcal{N}$ and $\mathcal{C}$ are convex sets, the condition can be tested numerically. By some algebraic manipulations, the following proposition transforms the verification of $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ to a LP problem.

Proposition 2. *Let $\mathbf{A} \in \mathbb{R}^{(N-K)\times(N-K)}$ be a submatrix of $\mathbf{J}(\boldsymbol{\gamma}^0)$ with row indices and column indices from $\mathcal{I}$. Let $\mathbf{C} \in \mathbb{R}^{K\times K}$ be a submatrix of $\mathbf{J}(\boldsymbol{\gamma}^0)$ with row indices and column indices from $\mathcal{I}^c$, where $\mathcal{I}^c$ is the complement of $\mathcal{I}$ with respect to $\{1, 2, \ldots, N\}$. Let $\mathbf{B} \in \mathbb{R}^{(N-K)\times K}$ be a submatrix of $\mathbf{J}(\boldsymbol{\gamma}^0)$ with row indices from $\mathcal{I}$ and column indices from $\mathcal{I}^c$. If $\mathbf{C}$ is invertible (which is usually the case in the considered problem), then the condition $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ in Theorem 1 is equivalent to the following feasibility problem*

$$\text{find} \quad \mathbf{x} \tag{12a}$$

$$\text{subject to} \quad (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)\mathbf{x} > \mathbf{0}, \tag{12b}$$

where vector $\mathbf{x} \in \mathbb{R}^{N-K}$.

Proposition 2 shows that if there exists a vector $\mathbf{x}$ in $\mathbb{R}^{N-K}$ such that $(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)\mathbf{x}$ lies in the positive orthant, then $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ holds. Note that the feasibility problem in (12) depends on the matrix $(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)$ only. The class of matrices that satisfies the constraint in (12b) is referred to as $\mathcal{M}^+$, which is introduced in [15] in the study of NNLS, and also used in [11] for the performance analysis via NNLS. Note that in this paper, we do not formulate the estimation of $\boldsymbol{\gamma}^0$ as an NNLS problem. Interestingly, the notion of $\mathcal{M}^+$ still appears (likely due to the non-negative constraint on $\boldsymbol{\gamma}$).

By solving (12) for different $\mathbf{J}(\boldsymbol{\gamma}^0)$ under various setups of N, L, K, σ_w^2, and randomly generated $\mathbf{S}$ and $\boldsymbol{\gamma}^0$, the necessary condition in Theorem 1 can be efficiently tested. Numerically we can identify a region in the space of N, L, K, σ_w^2 such that $\hat{\boldsymbol{\gamma}}$ can approach $\boldsymbol{\gamma}^0$ in the large M limit.

It is worth mentioning that an analytic form of the scaling law on N, L, K has been recently derived in [11] for the covariance based approach using the NNLS formulation and a binary constrained version of the MLE formulation. We emphasize that the phase transition analysis of Theorem 1 is for the general MLE formulation. Moreover, the condition in Theorem 1 provides a precise criterion for any given N, L, K, σ_w^2, sequence matrix $\mathbf{S}$, and $\boldsymbol{\gamma}^0$, whereas the scaling law of N, L, K presented in [11] is asymptotic in showing that the number of the identifiable active devices, K, is in the order of $\mathcal{O}(L^2)$, up to a logarithmic factor.

B. Distribution of Estimation Error $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0$

We now assume that the estimate $\hat{\boldsymbol{\gamma}}$ is consistent asymptotically as $M \to \infty$, and aim to characterize the distribution of the estimation error $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0$ for finite M. We achieve this by first characterizing the asymptotic distribution of $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0)$ in the regime $M \to \infty$, based on which we then obtain an approximated distribution of $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0$ for finite M. As mentioned before, $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0)$ does not tend to a Gaussian distribution due to the non-negative boundary. The following result considers the fact that $\boldsymbol{\gamma}^0$ is a boundary point, and characterizes the asymptotic distribution of $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0)$ via a QP problem. The result extends our previous work [9] with non-singular Fisher information matrix, which is based on [16], to the case of general Fisher information matrix.

Theorem 2. *Let $\mathbf{x} \in \mathbb{R}^{N\times 1}$ be a random vector sampled from the multivariate Gaussian distribution $\mathcal{N}\left(\mathbf{0}, M\mathbf{J}^\dagger(\boldsymbol{\gamma}^0)\right)$, where $\dagger$ denotes Moore-Penrose inverse. Let $\boldsymbol{\mu} \in \mathbb{R}^{N\times 1}$ be the solution to the following constrained QP*

$$\underset{\boldsymbol{\mu}}{\text{minimize}} \quad \frac{1}{M}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{J}(\boldsymbol{\gamma}^0)(\mathbf{x} - \boldsymbol{\mu}) \tag{13a}$$

$$\text{subject to} \quad \boldsymbol{\mu} \in \mathcal{C}, \tag{13b}$$

where $\mathcal{C}$ is defined in Theorem 1. Then $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0)$ has asymptotically the same distribution as $\boldsymbol{\mu}$ as $M \to \infty$.

Note that $\boldsymbol{\mu}$ is random due to the randomness of $\mathbf{x}$. Based on $\boldsymbol{\mu}$, the distribution of the estimation error $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0$ for finite M can be approximated from $M^{-\frac{1}{2}}\boldsymbol{\mu}$. Note that $\mathbf{x}$ is drawn from a degenerate multivariate Gaussian distribution if $\mathbf{J}^\dagger(\boldsymbol{\gamma}^0)$ is singular. An explanation of the QP in Theorem 2 is that we draw a sample $\mathbf{x}$ from the Gaussian distribution specified by the Fisher information matrix, and project the sample to the cone $\mathcal{C}$ such that the estimation error satisfies the constraints enforced by the fact that $\boldsymbol{\gamma}^0$ is on the boundary.

Since QP does not admit a closed-form solution in general, it is difficult to characterize the distribution analytically. However, Theorem 2 is still useful in the sense that it reveals the connection between the Fisher information matrix and the error distribution, and by solving (2) the distribution can be obtained numerically.

V. Numerical Results

Consider an mMTC system with one cell of radius 1000m. Assume that all devices are in the cell-edge for simplicity. The power of the background noise is -169dBm/Hz over 10 MHz, and the transmit power of each device is 23dBm.

We numerically test the necessary condition in Theorem 1 under a variety of choices of L and K, given $N = 1000, 1500$, or 2000. We draw the region of L, K in which the necessary condition is satisfied for each N. It is worth mentioning that we fix σ_w^2, whose value depends on the signal-to-noise ratio. We are interested in the case $L^2 < N$ where $\mathbf{J}(\boldsymbol{\gamma}^0)$ is singular. Further, since the estimation of K active devices is based on effectively $\mathcal{O}(L^2)$ observations of the covariance matrix, we plot L^2/N versus K/N in Fig. 1. Given L, K, we generate $\mathbf{J}(\boldsymbol{\gamma}^0)$ based on random $\mathbf{S}$ and $\boldsymbol{\gamma}^0$, and identify the

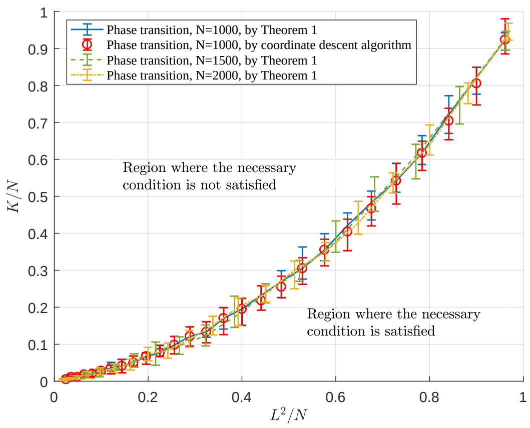

Fig. 1. Phase transition of the covariance based method with the MLE. The curve is based on 100 realizations of $\mathbf{S}$ and $\boldsymbol{\gamma}^0$ for each K, L. Error bars indicates the parameter range beyond which either all 100 realizations or 0 realizations satisfy the condition.

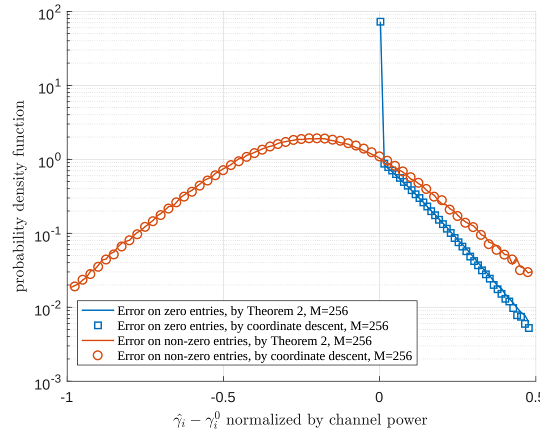

Fig. 2. Probability density functions (PDFs) of the error on the zero entries and the error on the non-zero entries.

region where condition can/cannot be satisfied. To validate the prediction by Theorem 1, we also run the coordinate descent algorithm to solve the MLE problem in (4) for $N = 1000$ in the large M limit, i.e, with ideal sample covariance matrix, to empirically plot a phase transition curve. We first observe from Fig. 1 that the curves with different N overlap, implying that the phase transition depends on N, L, K via the ratios $L^2/N, K/N$. We also observe that the curves by Theorem 1 and by the coordinate descent with $N = 1000$ match well, indicating that the necessary condition is sufficient for this example.

In Fig. 2, we validate the approximated distribution of $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0$ with $M = 256$ obtained from Theorem 2, by comparing it with the result from running coordinate decent to solve (4). We set $N = 1000$, $K = 50$, and $L = 20$ ($L^2/N = 0.4, K/N = 0.05$). For simplicity, we treat each coordinate of $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0$ as independent, and plot the empirical distribution of the coordinate-wise error. We consider two types of coordinates depending on if the true value on that coordinate is zero, and plot their distributions separately. We observe that the curves by Theorem 2 match those by solving (4) with coordinate descent in both cases. We also observe that there is a point mass in the distribution of the error for the zero entries. This is the probability that the inactive devices are correctly identified at finite $M = 256$.

VI. Conclusions

This paper studies the MLE approach to the pilot based device activity detection problem for mMTC with massive MIMO. We analyze the asymptotic performance of the covariance based MLE problem by studying its associated Fisher information matrix. This paper proposes a necessary condition on the Fisher information matrix such that the estimation error tends to zero in the massive MIMO regime, from which a phase transition analysis is obtained. The distribution of the estimation error is also analyzed.

References

[1] L. Liu, E. G. Larsson, W. Yu, P. Popovski, Č. Stefanović, 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 Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sept. 2018.

[2] K. Senel and E. G. Larsson, "Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6164–6175, Dec. 2018.

[3] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

[4] L. Liu and W. Yu, "Massive connectivity with massive MIMO —Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, June 2018.

[5] 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)*, June 2018, pp. 381–385.

[6] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June 2017, pp. 2523–2527.

[7] V. K. Amalladinne, A. Vem, D. K. Soma, K. R. Narayanan, and J.-F. Chamberland, "A coupled compressive sensing scheme for uncoordinated multiple access," 2018. [Online]. Available: http://arxiv.org/abs/1809.04745

[8] A. Fengler, G. Caire, P. Jung, and S. Haghighatshoar, "Massive MIMO unsourced random access," 2019. [Online]. Available: http://arxiv.org/abs/1901.00828

[9] Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Covariance based joint activity and data detection for massive random access with massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019.

[10] D. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.

[11] 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," 2019. [Online]. Available: http://arxiv.org/abs/1910.11266

[12] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, July 2007.

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

[14] T. J. Rothenberg, "Identification in parametric models," *Econometrica*, vol. 39, no. 3, pp. 577–591, May 1971.

[15] A. M. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4813–4820, Oct. 2008.

[16] S. G. Selfand and K.-Y. Liang, "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions," *J. Am. Stat. Assoc.*, vol. 82, no. 398, pp. 605–610, 1987.