Abstract—We propose a joint channel estimation and signal detection approach for the uplink non-orthogonal multiple access (NOMA) using unsupervised machine learning. We apply a Gaussian mixture model (GMM) to cluster the received signals, and accordingly optimize the decision regions to enhance the symbol error rate (SER) performance. We show that, when the received powers of the users are sufficiently different, the proposed clustering-based approach achieves an SER performance on a par with that of the conventional maximum-likelihood detector (MLD) with full channel state information (CSI). We study the tradeoff between the accuracy of the proposed approach and the blocklength, as the accuracy of the utilized clustering algorithm depends on the number of symbols available at the receiver. We provide a comprehensive performance analysis of the proposed approach and derive a theoretical bound on its SER performance. Our simulation results corroborate the effectiveness of the proposed approach and verify that the calculated theoretical bound can predict the SER performance of the proposed approach well. We further explore the application of the proposed approach to a practical grant-free NOMA scenario, and show that its performance is very close to that of the optimal MLD with full CSI, which usually requires long pilot sequences.

Index Terms—Cluster analysis, GMM, joint detection and estimation, massive IoT, NOMA, unsupervised machine learning.

I. INTRODUCTION

The fifth generation (5 G) of mobile standards has introduced two main service categories, namely, massive machine-type communications (mMTC) and ultra-reliable low-latency communications (URLLC) [1], alongside the enhanced mobile broadband (eMBB) that has been the primary focus of all previous generations. The main emphasis of mMTC is on the handling of a huge number of low-throughput, delay-tolerant, energy-efficient, and low-cost devices, which are usually used in Internet of things (IoT) applications. A recent forecast estimates that the number of enterprise and automotive IoT devices will grow from 5.8 billion in 2020 to 41.6 billion in 2025 generating 79.4 zettabytes of data per annum [2]. While 5 G is expected to serve many IoT applications, major breakthroughs in designing communication protocols and radio resource management techniques are required to serve applications with a diverse range of requirements in terms of data rate, reliability, availability, end-to-end latency, energy efficiency, security, and privacy [3].

A. Related Work

Non-orthogonal multiple access (NOMA) schemes have gained significant attention in the last decade as an enabler for mMTC [4]. NOMA, either power-domain or code-domain [5], allows multiple users to share the same radio resources, leading to higher spectral efficiency. While in power-domain NOMA users transmit with different power levels depending on their channel conditions, code-domain NOMA relies on assigning a unique code to each user. Recent studies have explored NOMA-based approaches to safeguard secrecy, especially in UAV-assisted applications and satellite communications [6], [7]. Such works highlight the inherent potential of NOMA not only as a tool for optimizing performance but also as a mechanism to enhance secure transmissions. Drawing from these insights, it becomes evident that NOMA can serve as a robust and versatile technique, capable of addressing both performance and security challenges in next-generation communication paradigms. Not only does NOMA allow multiple devices to share the same radio resources, but it can also reduce the signaling overhead and latency by allowing grant-free uplink connection [8]. Incorporation of NOMA with the grant-free access that is a lightweight random access protocol is considered to be a key enabler of massive connectivity in IoT [4].

In the massive IoT served by 5 G and beyond, the communication network will be highly complex and dynamic with numerous possibly contending design requirements [9], [10]. Machine learning can be effectively used to enhance network intelligence and enable self-adaptability in an efficient and timely manner [11]. Machine learning can be used to render effective and efficient services in different network layers. In the physical layer, clustering methods, such as $k$-means and Gaussian mixture model (GMM) can be utilized for channel estimation and signal detection while convolutional neural network algorithms can be applied for channel decoding [12]. To increase reliability, powerful supervised machine learning algorithms such as deep neural networks can be employed at the data link layer for user scheduling. Moreover, reinforcement learning can be utilized at
the network layer to improve network robustness and service continuity [13], [14].

Channel estimation schemes can be mainly categorized into training-based [15], blind [16], and semi-blind [17]. Training-based schemes use long training sequences to obtain the channel state information (CSI) with low complexity but at the expense of decreased throughput [15]. In contrast, by using the properties of the transmitted signal, blind channel estimation schemes estimate the channel without any training symbols [16]. While these schemes are bandwidth efficient, they are more complex and less accurate in comparison with training-based schemes. To strike a balance between throughput, accuracy, and complexity, semi-blind schemes combine the merits of both training-based and blind schemes [18]. Recently, there has been a growing interest in joint channel estimation and signal detection for grant-free NOMA [19], [20], [21]. By representing the joint problem as a compressed sensing problem, the authors of [19], [22], [23] utilize the sparsity of the pilot (training) symbols to reduce their number. However, in massive IoT, where the packets are usually small, even a few training symbols can lead to a major efficiency loss [24].

Unsupervised machine learning algorithms can be used for joint channel estimation and signal detection [25]. For an uplink NOMA scenario, the probability density function (PDF) of the received signals can be modelled by a GMM and the maximum-likelihood estimation (MLE) can be used to estimate the model parameters. While the MLE is asymptotically efficient, it may be intractable or biased, especially with a small sample size [26], [27]. In such cases, the expectation-maximization (EM) algorithm is useful for estimating the model parameters [28], [29] in an iterative manner. EM is numerically stable and increases the likelihood function with each iteration [30]. The global convergence of the EM algorithm fitting a GMM model consisting of two Gaussian distributions with similar densities is guaranteed, as long as it is initialized in the neighborhood of the desired solution [31]. For higher-order mixture models, if the GMM centers (the means of the Gaussian components) are well separated, EM converges locally to the global optimum [32], [33]. The lower and upper bounds for the number of samples required to estimate the parameters of a GMM with well-separated centers given any desired accuracy is calculated in [34], [35].

B. Motivations and Contributions

In our recent conference article [24], the main idea of using clustering algorithms for joint channel estimation and signal detection is presented for an uplink NOMA scenario with quadrature phase shift keying (QPSK) modulation. Using GMM clustering, we can cluster the received signals into $M$ clusters, where $M$ is the modulation order, in each stage of successive interference cancellation (SIC). The channel gain $|h|$ can then be estimated using the coordinates of each cluster centroid. However, this does not provide any information about the phase of the channel. In [24], we considered sending a few pilot symbols to determine the demapping. However, we did not take into account the pilot length and its impact on the throughput of the system. The results in [24] show that, by using GMM clustering with only a few pilot symbols, the receiver can effectively estimate the users’ channels and detect the signals with a SER approaching that of the optimal receiver, that is, the maximum likelihood detection (MLD) with full CSI. The primary focus of our study [25] was on the utilisation of clustering algorithms and rotational-invariant codes for the purpose of joint channel estimation and signal detection in an uplink NOMA scenario. Nevertheless, the aforementioned study lacks a comprehensive mathematical framework for accurately characterising the performance of the error rate, convergence, and estimation error boundaries. Additionally, the study conducted by [25] does not provide a comprehensive analysis comparing the suggested algorithm with other existing approaches that utilise training-based and semi-blind channel estimation techniques. Furthermore, the aforementioned study [25] lacks a comprehensive analysis of the practical applications and the effects of algorithmic parameters.

In this article, we propose a generalized GMM-based joint channel estimation and signal detection algorithm for NOMA. We provide a theoretical analysis of the proposed approach to prove its convergence and derive a closed-form expression for its SER. We also study the tradeoff between pilot length and SER offered by the proposed approach, and further explore its application in a grant-free NOMA scenario suitable for massive IoT systems. Our main contributions in this article are summarized below.

1) We first propose a generalized GMM-based joint channel estimation and signal detection algorithm for NOMA. We provide a mathematical analysis to characterize the error rate performance of the proposed algorithm. We prove the convergence of the proposed GMM clustering algorithm and derive an upper bound on its error. Accordingly, we calculate a bound on the channel estimation error, in particular, the gain and phase errors, to derive a closed-from expression for the SER. We verify the accuracy of the derived bound via simulations. Furthermore, we show that when the powers of the signals received from the users are sufficiently different, the proposed clustering-based approach with no CSI at the receiver achieves the same SER performance as the conventional MLD receiver with full CSI. We also explore the tradeoff between accuracy and blocklength as the accuracy of clustering algorithm depends on the number of data points (symbols) available at the receiver. We also show that the proposed approach can be easily extended for NOMA with higher-order modulations.

2) We provide a comprehensive comparison between the proposed approach and some existing related ones that use training-based or semi-blind channel estimation. Training-based channel estimation usually requires a large number of pilot symbols to accurately estimate the CSI. We show that our proposed approach considerably outperforms the training-based and semi-blind channel estimation techniques when only a few pilot symbols are available. We also show, the proposed approach requires a significantly lower number of pilot symbols to achieve the benchmark MLD performance compared with existing training-based and semi-blind channel estimation techniques.

3) We explore the application of the proposed GMM-based joint channel estimation and signal detection algorithm in a practical grant-free NOMA scenario, where the receiver does not have any knowledge of the number of active users or the channel conditions. The results show that the proposed approach almost achieves the MLD performance using a small number of pilot symbols. This is of particular advantage for massive IoT applications, where the users’ activity is usually arbitrary and the receiver cannot estimate the channels of all users promptly.
C. Paper Organization and Notations

The rest of the article is organized as follows. We present the system model and provide some relevant preliminary information in Section II. We describe our proposed GMM-clustering-based approach for joint channel estimation and signal detection in Section III. In Section IV, we characterize theoretical bounds on the SER of the proposed approach. We provide extensive numerical results to evaluate the performance of the proposed approach in Section V. The application of the proposed approach in grant-free massive IoT is presented in Section VI. Finally, Section VII concludes the article.

Notations: Matrices, vectors, and scalars are denoted by uppercase boldface, lowercase boldface, and lowercase letters, e.g., $\mathbf{A}$, $\mathbf{a}$, and $a$, respectively. The transpose, inverse and norm of $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^{-1}$, and $||\mathbf{A}||$, respectively.

II. System Model and Preliminaries

A. Channel Model

We consider a cellular uplink NOMA scenario in which $K$ active users simultaneously transmit packets of length $N$ symbols to a base station (BS). Let $x_i$ denote the signal transmitted by user $i$, which is drawn from the signal constellation $\mathcal{S} = \{s_1, s_2, \ldots, s_M\}$ with $M = |\mathcal{S}|$ being the modulation order. The received signal at the BS, denoted by $y$, is given by

$$y = \sum_{i=1}^{K} h_i \sqrt{P_i} x_i + w,$$

(1)

where $P_i$ is the transmit power of user $i$ and $w \sim \mathcal{CN}(0, 1)$ is additive white Gaussian noise (AWGN). The gain of the channel between user $i$ and the BS is denoted by $h_i$, which includes both small-scale and large-scale fading, i.e., $h_i = g_i \sqrt{\frac{d_{0}}{d_i}} r_i \chi$, where $d_0$ and $r_0$ are the reference path-loss and reference distance, respectively, $\chi$ is the path-loss exponent, $r_i$ is the distance between user $i$ and the BS, $\chi$ is the large-scale shadowing modelled by a log-normal distribution with zero mean and variance $\sigma$ dB, and $g_i$ is the small-scale fading modelled by the Rayleigh, Rician, Nakagami, or any other distribution. The signal-to-noise ratio (SNR) of user $i$ at the BS is given by $\gamma_i = P_i h_i^2$. We assume that $h_i$ remains constant for the duration of one packet ($N$ symbols), which is a valid assumption for short packets, especially in mMTC applications [36], [37].

Given the user channel gains and knowing that all users utilize the same modulation with constellation set $\mathcal{S}$, the probability distribution of the received signals at the BS can be expressed as a mixture of Gaussian distributions.\footnote{One can easily show that this is also valid when users use different modulations, and the assumption is made only for a better representation in the article.} Let $\mathbf{h} = [h_1, h_2, \ldots, h_K]$, $\mathbf{p} = [P_1, P_2, \ldots, P_K]$, and $\mathbf{x} = [x_1, x_2, \ldots, x_K]$ denote the vectors of channel gains, transmit powers, and user signals, respectively. Therefore, we have

$$p(y|\mathbf{h}, \mathbf{p}, \mathcal{S}) = \sum_{\mathbf{u} \in \mathcal{S}^K} p(\mathbf{u} = \mathbf{x}) p(y|\mathbf{h}, \mathbf{p}, \mathcal{S}, \mathbf{x} = \mathbf{u})$$

(2)

$$= \sum_{\mathbf{u} \in \mathcal{S}^K} \prod_{i=1}^{K} p(x_i = u_i) p(y|\mathbf{h}, \mathbf{p}, \mathcal{S}, \mathbf{x} = \mathbf{u})$$

(3)

where $u_i$ are the reference path-loss $\alpha$ and $\chi$ is the path-loss exponent, $r_i$ is the distance between user $i$ and the BS, $\chi$ is the large-scale shadowing modelled by a log-normal distribution with zero mean and variance $\sigma$ dB, and $g_i$ is the small-scale fading modelled by the Rayleigh, Rician, Nakagami, or any other distribution. The signal-to-noise ratio (SNR) of user $i$ at the BS is given by $\gamma_i = P_i h_i^2$. We assume that $h_i$ remains constant for the duration of one packet ($N$ symbols), which is a valid assumption for short packets, especially in mMTC applications [36], [37].

Given that the received signals are complex numbers, we denote $\mathbf{h}$ as a 2-dimensional multivariate Gaussian probability density function by $g(\mathbf{z}; \mu, \Sigma)$ where $\mu$ and $\Sigma$ are the mean vector and the covariance matrix, respectively, and express it as

$$g(\mathbf{z}; \mu, \Sigma) = \frac{1}{(2\pi)^{||\Sigma||}} \exp \left( -\frac{1}{2} (\mathbf{z} - \mu)^T \Sigma^{-1} (\mathbf{z} - \mu) \right).$$

(6)

In GMM clustering, the number of clusters is pre-determined and the data is assumed to be generated by a mixture of Gaussian distributions. A GMM parameterizes the mean, covariance,
and weight of each Gaussian distribution component. When a common \(M\)-ary modulation scheme is adopted by all users, there are \(M\) Gaussian distributions each with a nonnegative mixture weight \(\omega_j\) where \(j \in \{1, \ldots, M\}\) and \(\sum_{j=1}^{M} \omega_j = 1\). Accordingly, the underlying Gaussian mixture distribution can be written as a convex combination of \(M\) constituent Gaussian distributions (each representing a cluster), i.e.,

\[
p(z; \mu_1, \ldots, \mu_M, \Sigma_1, \ldots, \Sigma_M) = \sum_{j=1}^{M} \omega_j g_j(z; \mu_j, \Sigma_j). \tag{7}
\]

We are interested in estimating \(\mu_j, \Sigma_j\), and \(\omega_j\), \(j = 1, \ldots, M\), from the observed data. This can be done by maximizing the likelihood function (7) for all received signals. To this end, we utilize the EM algorithm [40] that is suitable for solving maximum-likelihood problems containing unobserved latent variables.

We define the corresponding log-likelihood function as

\[
l^{(t)}(\mu_1, \ldots, \mu_M, \Sigma_1, \ldots, \Sigma_M|z_1, \ldots, z_N) = \sum_{i=1}^{N} \left[ \sum_{j=1}^{M} \Delta^{(t)}_{i,j} \ln \left( \frac{\omega_j^{(t)} g_j(z_i; \mu_j^{(t)}, \Sigma_j^{(t)})}{\sum_{k=1}^{M} \omega_k^{(t)} g_k(z_i; \mu_k^{(t)}, \Sigma_k^{(t)})} \right) \right]. \tag{8}
\]

We evaluate this function only to verify the convergence of the EM algorithm. Let \(\Delta_{i,j}\) symbolize the association of the \(i\)th data point to the \(j\)th cluster represented by the \(j\)th Gaussian distribution. Therefore, we have

\[
\Delta_{i,j} = \begin{cases} 1 & \text{if } z_i \text{ belongs to cluster } g_j, \\ 0 & \text{otherwise}. \end{cases}
\]

It is clear that \(P(\Delta_{i,j} = 1) = \omega_j\) and \(P(\Delta_{i,j} = 0) = 1 - \omega_j\). However, both \(\Delta_{i,j}\) and \(\omega_j\) are unknown. In the \(t\)th iteration of the EM algorithm, we first estimate the so-called responsibility variable for each model \(j\) and every data point \(i\) as

\[
\hat{\gamma}^{(t)}_{i,j} = \frac{\omega_j^{(t-1)} g_j(z_i; \mu_j^{(t-1)}, \Sigma_j^{(t-1)})}{\sum_{k=1}^{M} \omega_k^{(t-1)} g_k(z_i; \mu_k^{(t-1)}, \Sigma_k^{(t-1)})}. \tag{9}
\]

We then assign each data point to its corresponding cluster. In particular, for each \(z_i\), we find \(m_i^{(t)} = \arg \max_j \hat{\gamma}^{(t)}_{i,j}\) and set

\[
\Delta^{(t)}_{i,j} = \begin{cases} 1 & \text{if } j = m_i^{(t)}, \\ 0 & \text{otherwise}. \end{cases}
\]

In the next step of the EM algorithm, we use the calculated responsibilities to update the mixture weight, mean, and variance for each cluster as follows

\[
\hat{\omega}_j^{(t)} = \frac{\sum_{i=1}^{N} \hat{\gamma}^{(t)}_{i,j}}{\sum_{i=1}^{N} \sum_{k=1}^{M} \hat{\gamma}^{(t)}_{i,k}}, \tag{10}
\]

\[
\hat{\mu}_j^{(t)} = \frac{\sum_{i=1}^{N} \hat{\gamma}^{(t)}_{i,j} z_i}{\sum_{i=1}^{N} \hat{\gamma}^{(t)}_{i,j}}, \tag{11}
\]

\[
\hat{\Sigma}_j^{(t)} = \frac{\sum_{i=1}^{N} \hat{\gamma}^{(t)}_{i,j} \left[ z_i - \hat{\mu}_j^{(t)} \right] \left[ z_i - \hat{\mu}_j^{(t)} \right]^T}{\sum_{i=1}^{N} \hat{\gamma}^{(t)}_{i,j}}. \tag{12}
\]

After enough iterations, the values of responsibility, mean, covariance, and mixture weight for each cluster converge as the EM algorithm is guaranteed to converge to a local optimum [29]. The number of iterations required for convergence mainly depends on the convergence criterion. In Fig. 1, we visualize the convergence of the mean estimates (cluster centroids) after running eight iterations of the EM algorithm.

There are several other prominent clustering algorithms, such as k-means, DBSCAN [41], OPTICS [42], and mean shift [43], that can also be used to cluster the received signals. However, as the noise has Gaussian distribution, the received signals at the BS naturally follow a mixture of Gaussian distributions (5). Hence, GMM is a good choice for our clustering problem. From a theoretical standpoint, given that the noise affecting the received signals is independent and identically distributed (i.i.d.) AWGN, the GMM clustering used in our proposed algorithm is equivalent to the well-known k-means clustering algorithm. However, in practise, due to the limited amount of observations, the noise covariance matrix is not a multiple of the identity matrix, which means that the noise affecting different received signals may be correlated or have different variances. As a result, GMM clustering is more accurate compared to k-means clustering since, unlike k-means, it does not assume the same covariance for all clusters but estimates them from the data.

III. CLUSTERING-BASED JOINT CHANNEL ESTIMATION AND SIGNAL DETECTION

To better understand the proposed joint channel estimation and signal detection approach, we first discuss Fig. 2 that illustrates the signals collected at the receiver in the I-Q plane for a two-user NOMA communication system when both users employ the QPSK modulation. We assume that \(|h_1| \geq |h_2|\) without losing generality. Fading causes signal amplification and a rotation with respect to the original constellation. When the received powers for the two users are significantly different, the clusters are distinct from each other. Therefore, the phase

2We consider our approach to be a joint operation due to the inherent interdependence between the channel estimation and signal detection in the proposed algorithm, as the estimated channels are prerequisite for correct detection and removal of each user’s signal, highlighting the mutual reliance of these two processes.
Algorithm 1: Joint Channel Estimation and Signal Detection Using GMM Clustering.

**Input:** number of users $K$, received signal $y$, modulation order $M$, signal constellation $S$, and convergence threshold $\epsilon$

**Output:** estimated signals of the users

1. Set $\omega_j = \frac{1}{M}$, $j = 1, \ldots, M$
2. for user $u = 1$ to $K$ do
   3. Initialize $\mu^{(0)}_u$ and $\Sigma^{(0)}_u$ by dividing the received signals on the reference coordinate system into $M$ equal sections, select one point in each section as the initial mean and set the initial covariance to one
   4. Calculate $\gamma^{(0)}_{i,j}$ according to (9)
   5. Calculate log-likelihood function according to (8)
   6. Set $t = 1$
   7. while $l^{(t)} - l^{(t-1)} \geq \epsilon$ do
      8. Update $\mu^{(t)}_u$ and $\Sigma^{(t)}_u$ using (11) and (12)
      9. Update $\gamma^{(t)}_{i,j}$ according to (9)
     10. Update log-likelihood function according to (8)
   11. end
   12. Return $\mu = \bar{\mu}^{(t)}$ and $\Sigma = \bar{\Sigma}^{(t)}$
   13. Calculate the phase of each cluster centroid: $\phi_i = \tan^{-1} \left( \frac{\text{Im}(\mu_i)}{\text{Re}(\mu_i)} \right)$
   14. Calculate the average phase as $\bar{\phi} = \frac{1}{M} \sum_{i=1}^{M} \phi_i$ and channel amplitude $|\bar{h}_u| = \frac{1}{M} \sum_{i=1}^{M} |\mu_i|$
   15. Update the decision boundaries based on the $\phi$ and use the pilot symbol to map each cluster into the mapping bits
   16. Demodulate the signal to symbols: $\hat{x}_u = \text{demod}(y)$
   17. Re-modulate this user’s signal and multiply by the estimated channel gain and subtract it from superimposed received signal: $y \leftarrow y - |\bar{h}_u|e^{j\bar{\phi}} \hat{x}_u$
18. end

and amplitude of the channels can be estimated accurately. However, when channel fading and noise cause the clusters to overlap, estimation of the channels and detection of the signals is more challenging and inevitably less accurate. In what follows, we propose an effective algorithm to jointly estimate the user channels by clustering the received signals and detect the user signals.

At the BS, SIC is used to recover the multiplexed user signals from the received superimposed signals in the decreasing order of their received powers. Initially, the receiver detects the signal of the strongest user that is the user with the highest received power. It then reconstructs and removes it from the received signal. Afterwards, it detects the next strongest signal and so on. In order to implement SIC at the BS, we start by dividing the received signals (data points) into $M$ (four with QPSK) clusters, which represent the user 1’s signals. Next, we further split each of the $M$ clusters into $M$ smaller clusters representing the user 2’s signals and so on. Taking the received signals as our observed data, the considered joint channel estimation and signal detection problem boils down to estimating the unknown latent parameters of the assumed Gaussian mixture distribution in (5).

A. The Proposed Algorithm

We summarize the proposed algorithm for joint channel estimation and signal detection in Algorithm 1. In the proposed algorithm, at each stage of SIC, we estimate the parameters of only $M$ Gaussian distributions. This helps with managing the computational complexity.

Given the modulation order $M$, we fix the mixture weight of each cluster as $\omega_j = \frac{1}{M}$. Next, starting from the first user, we split the received data into $M$ clusters and select one point in each cluster as the initial mean and set the initial variance of each cluster to one. Then, we calculate the responsibility and log-likelihood function values using (9) and (8), respectively. Afterwards, we calculate the associated cluster centroids and covariance matrices (lines 8 to 13 in Algorithm 1). We continue by evaluating the phase of each cluster. Considering that we are using the QPSK modulation, the phase difference between any two adjacent clusters is $\frac{\pi}{2}$. However, due to noise, the phase of each centroid might differ from $\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4},$ or $\frac{7\pi}{4}$. To minimize the effect of phase rotation, we average the phase difference between the centroid of each cluster and its expected value (line 14 of Algorithm 1), and update the decision boundaries accordingly to the average phase rotation. Finally, we apply SIC and repeat the algorithm for the next strongest user.

It is important to note that by using SIC, we assume the first $M$ clusters to be Gaussian each consisting of $M$ subclusters, although, the clusters corresponding to the strongest user are not strictly Gaussian in practice. However, roughly speaking, the components of the GMM represent clusters that are of similar shapes. Thus, each cluster is implicitly assumed to follow a Gaussian distribution. For cases where this assumption is overly unrealistic, a natural alternative is to assume that each cluster is also a mixture of normally-distributed subclusters [44].

B. Symbol-to-Bit Demapping

Using GMM clustering, we can separate the received signals into $M$ clusters in each stage of SIC. We can then estimate the channel gain $|h|$ from the coordinates of the cluster centroids (line 14 of Algorithm 1). However, this does not inform us about the exact phase of the channel. In particular, assuming that the user sends QPSK symbols, according to Fig. 3, one
can consider four different but equally probable choices for the phase of the channel, i.e., $\angle h \in \{\theta - \pi/4, \theta - 3\pi/4, \theta + \pi/4, \theta + 3\pi/4\}$ where $-\pi/4 < \theta < \pi/4$ is the phase of the cluster that resides in the angular region of $[-\pi/4, \pi/4]$. This implies that we cannot demap the signals to the symbols.

To overcome this issue, we send a few pilot symbols to determine the demapping. Determining the mapping with only one symbol is unambiguous when the SNR is not excessively low. The users’ pilot symbols are distinct from one another and in the scenario when there are two users, each with two pilot symbols, we consider Pilot$_{u_1} = [s_1, s_3]$ and Pilot$_{u_2} = [s_1, s_3]$. Furthermore, the BS receives superimposed modulated signals. For a robust and accurate performance, two symbols per user can be sufficient. We use a minimum mean square error (MMSE) method to estimate the phase rotation due to the channel. The MMSE estimate is given as

$$\hat{h}_{u,\text{MMSE}} = (x_{u,p}^H x_{u,p} + 1)^{-1} x_{u,p}^H y_p \tag{13}$$

where $x_{u,p}$ is the transmitted pilot of user $u$ and $y_p$ is the superimposed signal received at the BS. The phase of the channel is determined to be the one from $\{\theta - \pi/4, \theta - 3\pi/4, \theta + \pi/4, \theta + 3\pi/4\}$ that is closest to $\hat{h}_{u,\text{MMSE}}$. Using this method, the exact phase rotation due to the channel is not required. As we will show later, this is another advantage of our proposed approach over the conventional channel estimation techniques. Referring to Fig. 3, for two-user NOMA, we may assume that, in the first time slot, both users send $s_1$ while, in the second time slot, the first user transmits $s_2$ and the second user sends $s_3$ as the pilot symbols.

In our proposed algorithm, the receiver initially lacks any prior knowledge of the channel coefficients. It estimates these coefficients by leveraging the combined information derived from the GMM clustering and the pilot symbols. After estimating the channel gain through GMM clustering and determining the phase rotation using pilot symbols for demapping, the information of the stronger user is decoded. Subsequently, the user’s signal is re-modulated and removed from the superimposed signal. This process is repeated for the subsequent users.

IV. THEORETICAL ANALYSIS

In this section, we analyze the performance of the proposed algorithm. In each iteration of EM, the mean, variance, and mixture weight of each cluster are updated. We refer to the difference between the mean of each cluster estimated by EM and its corresponding exact value as the EM error. We start by presenting a theorem that gives an upper-bound on the EM error. Then, using this bound, we present a mathematical model to predict the SER of our proposed GMM-clustering-based joint channel estimation and signal detection algorithm.

Let $R_{\min}$ and $R_{\max}$ denote the distance between the closest cluster centroids and the distance between the farthest cluster centroids, respectively, as shown in Fig. 4. The following theorem characterizes an upper-bound on the EM error.

**Theorem 1:** Assume $d$-dimensional received signals and $M$ isotropic Gaussian distributions with mixture weights $\omega_j$, $j \in \{1, \ldots, M\}$. Let $\mu_i^\star$ denote the true mean of cluster $i$ and $\hat{\mu}_i^{(t)}$ represent the estimated mean of the $i$-th cluster after $t$ iterations of EM. Suppose $\kappa = \min \{\omega_j\}$, $R_{\min} \geq C_0 \sqrt{\min\{d, M\}}$, and the initial iterate $\mu^{(0)}$ satisfies

$$\max_{i \in [M]} ||\mu^{(0)}_i - \mu_i^\star||_2 \leq \frac{R_{\min}}{2}$$

$$- \frac{C_1}{2} \sqrt{\min\{d, M\}} \log \left( \max \left\{ \frac{M}{\kappa^2}, R_{\max}, \min\{d, M\} \right\} \right)$$

where $[M] = \{1, 2, \ldots, M\}$ and $C_0, C_1 > 0$ are universal constants. For a sufficiently large sample size $N$ such that that

$$\frac{\log(N)}{N} \leq \min \left\{ \frac{\kappa^2}{144C_2Md}, \frac{\kappa^2 \max_{i \in [M]} ||\mu^{(0)}_i - \mu^\star_i||_2^2}{9C_3R_{\max}^2Md} \right\}$$

where $C_2, C_3 > 0$ are universal constants and

$$\hat{C}_2 = C_2 \log \left( M \left( 2R_{\max} + \sqrt{d} \right) \right)$$

$$\hat{C}_3 = C_3 \log \left( M \left( 3R_{\max}^2 + \sqrt{d} \right) \right),$$

the subsequent EM iterates $\{\mu_i^{(t)}\}_{t=1}^\infty$ satisfy

$$\max_{i \in [M]} ||\mu_i^{(t)} - \mu_i^\star||_2 \leq \frac{1}{2} \max_{i \in [M]} ||\mu_i^{(0)} - \mu_i^\star||_2$$

$$+ \frac{3R_{\max}}{\kappa} \sqrt{\hat{C}_2Md \log(N) / N}$$

with probability at least $1 - \frac{2M}{N}$.

**Proof:** See Appendix.

Theorem 1 states that the EM error is bounded. As the iteration number, $t$, increases, the first term on the right-hand side of (16) converges to zero and the second term dominates. The following remark characterizes this bound.

**Remark 1:** When $t \to \infty$, the EM error characterized in Theorem 1 is upper bounded by

$$\max_{i \in [M]} ||\mu_i^{(\infty)} - \mu_i^\star||_2 \leq \frac{3R_{\max}}{\kappa} \sqrt{\hat{C}_3Md \log(N) / N}.$$
This shows that EM converges to points within balls of radius \( \frac{3R_{\max}}{\kappa} \sqrt{C_3M d \log(N) / N} \) around the true centers. Fig. 4 illustrates the essence of Theorem 1 when \( M = 4 \). As seen in the figure, there is a circle of radius \( \frac{3R_{\max}}{\kappa} \sqrt{C_3M d \log(N) / N} \) around each true center and the EM converges to points within these circles. Moreover, the EM error converges to zero when the sample size, \( N \), is arbitrarily large. This follows directly from (17) by taking \( N \) to infinity, noting that \( \hat{C}_3 \) is constant.

Remark 2: The EM error characterized in Theorem 1 converges to zero when \( t \to \infty \) and \( N \to \infty \).

The following lemma provides an approximation for the SER of the proposed GMM-clustering-based joint channel estimation and signal detection algorithm when \( M = 4 \), i.e., when QPSK modulation is used.

Lemma 1: For a QPSK-modulated signal, i.e., \( M = 4 \) and \( d = 2 \), at SNR \( \gamma \) with \( N \) samples, the proposed algorithm achieves an SER approximated by

\[
P_{\text{SER}} \approx Q \left( \sqrt{2\gamma \sin \left( \frac{\pi}{4} - \varphi \right)} \right) + Q \left( \sqrt{2\gamma \sin \left( \frac{\pi}{4} + \varphi \right)} \right)
\]

(18)

where

\[
\varphi = \tan^{-1} \left( 6M \sqrt{C_3M d \log(N) / N} \right)
\]

(19)

and \( Q(z) = \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{-d^2} dv \).

Proof: Considering Remark 1, each estimated cluster centroid has a distance of at most \( \frac{3R_{\max}}{\kappa} \sqrt{C_3M d \log(N) / N} \) to the true centroid when \( t \to \infty \). This means for QPSK-modulated signals (\( M = 4 \)), as shown in Fig. 4, there is a phase mismatch between the final cluster centroids and their corresponding true centroids given by

\[
\varphi = \tan^{-1} \left( \frac{3R_{\max}}{\kappa} \sqrt{C_3M d \log(N) / N} / \frac{R_{\max}}{2} \right)
\]

\[= \tan^{-1} \left( 6M \sqrt{C_3M d \log(N) / N} \right)
\]

(20)

where we use \( \kappa = 1 / M \). Therefore, the detection of the QPSK signals is with a rotated phase reference of \( \varphi \). In [45], it is shown that the SER for a QPSK signal with a noisy phase reference \( -\pi / 4 \leq \varphi \leq \pi / 4 \) is given by (18). This completes the proof.

Remark 2 states that, when the number of samples \( N \), i.e., the number of signals received at the BS, is sufficiently large, the EM error converges to zero. This means that the GMM-clustering-based joint channel estimation and signal detection algorithm will find the true locations of the cluster centroids (shown by black dots in Fig. 4). Thus, the mismatch in the phase reference \( \varphi \) will be zero (Lemma 1) and \( P_{\text{SER}} \approx 2Q(\sqrt{\gamma}) \). Therefore, the SER of the proposed algorithm will be the same as that of the MLD with full CSI.

We can further extend the results in Lemma 1 to approximate the SER of the proposed algorithm for the two-user NOMA scenario.

Lemma 2: For a two-user NOMA scenario with \( N \) QPSK-modulated signals where the SNR of users 1 and 2 are \( \gamma_1 \) and \( \gamma_2 \), respectively, the proposed algorithm yields per-user SERs approximated by

\[
P_{\text{SER},1} \approx \frac{1}{4}Q \left( \sqrt{2\gamma_1 + \sqrt{2\gamma_2}} \sin \left( \frac{\pi}{4} - \varphi_1 \right) \right)
\]

\[+ \frac{1}{4}Q \left( \sqrt{2\gamma_1 + \sqrt{2\gamma_2}} \sin \left( \frac{\pi}{4} + \varphi_1 \right) \right)
\]

\[+ \frac{1}{4}Q \left( \sqrt{2\gamma_1 - \sqrt{2\gamma_2}} \sin \left( \frac{\pi}{4} - \varphi_1 \right) \right)
\]

\[+ \frac{1}{4}Q \left( \sqrt{2\gamma_1 - \sqrt{2\gamma_2}} \sin \left( \frac{\pi}{4} + \varphi_1 \right) \right)
\]

(21)

\[
P_{\text{SER},2} \approx Q \left( \sqrt{2\gamma_2 \sin \left( \frac{\pi}{4} - \varphi_2 \right)} \right) + Q \left( \sqrt{2\gamma_2 \sin \left( \frac{\pi}{4} + \varphi_2 \right)} \right)
\]

(22)

where \( \varphi_i, i \in \{1, 2\} \), is given by

\[
\varphi_i = \tan^{-1} \left( 6M \sqrt{C_3^{(i)}M d \log(N) / N} \right)
\]

(23)

and \( C_3^{(i)} = C_3 \log(M(12\gamma_i + \sqrt{d})) \).

Proof: Authors of [46] characterize the SER of the QPSK constellation for the uplink NOMA scenario. It is shown that the SER for the stronger user, user 1 in our case, is approximated by

\[
Q \left( \sqrt{\gamma_1} + \sqrt{\gamma_2} \right) + Q \left( \sqrt{\gamma_1} - \sqrt{\gamma_2} \right)
\]

(24)

In the proposed approach, we have the phase mismatch \( \varphi_1 \) [see (23) in Lemma 1] between the final cluster centroids and the true centroids for user 1’s signals. Similar to [45], by incorporating this phase difference into each component of (24), (21) can be easily derived. After detecting user 1’s signal, we apply SIC and subtract the detected signal from the received signal. Detecting user 2’s signal is then straightforward as it is a noisy QPSK signal with the interference of the other user canceled. Therefore, using Lemma 1, the SER of user 2 can be approximated by (22). This completes the proof.

Note that the convergence of GMM as per Theorem 1 is local since we assume that the EM algorithm is initialized in the neighborhood of the true centroids. As the pilots are in close proximity to the true cluster centroids, they can be utilised to initialise the EM algorithm as they are at the centre of the decision boundaries. When the BS does not have the knowledge of the modulation scheme, other techniques such as the method of moments [47] can be used for the initialization.

Since the EM algorithm used for GMM clustering comprises two alternating steps of expectation and maximization, to determine its computational complexity, we consider the complexity of each step. For a general case, assume we have \( d \) dimensions, \( M \) clusters, and \( N \) samples. In the expectation step, we calculate the determinant and the inverse of the covariance matrix with the complexity of \( O(MN d^3) \). The maximizing stage entails calculating the mixture weight, mean, and covariance for each cluster with corresponding \( O(MN) \), \( O(MN) \), and \( O(M N d) \) complexity levels. Given the algorithm convergences after \( t \) iterations, the computational complexity of GMM clustering

\[A more accurate approximation can be obtained by considering the phase rotation for each user. However, the approximations in (21) and (22) are sufficiently accurate while finding more accurate approximations is beyond the scope of this article.\]
is \(O(tMN d^3)\), which is most importantly linear in \(N\). Considering the QPSK modulation, \(N = 50\) data points in a two-dimensional dataset, and convergence within three iterations, our algorithm requires around 4800 operations. For a packet length four times that of reference [48], the computational complexity of our proposed algorithm is similar to that of the deep-learning-based algorithm proposed in [48]. Although our algorithm is computationally more complex compared to the SIC-based joint ML detector using full CSI, it has the advantage of eliminating the requirement for full CSI. It efficiently performs joint channel estimation and signal detection using a minimal number of pilot symbols. Furthermore, given a similar number of pilot symbols, our algorithm outperforms the MLD, offering significant advantages in scenarios with high numbers of devices. Since all computations are carried out at the BS, which generally has sufficient computing and energy resources, the mentioned computational complexity does not pose any challenge, particularly considering the substantial gains in throughput.

V. NUMERICAL RESULTS

In our simulations, we consider that two pilot symbols are sent for symbol-to-bit demapping, unless we specify otherwise. We also assume that the convergence threshold, \(\epsilon\), for the proposed GMM-clustering-based approach is set to 1, unless specified otherwise.

A. Single-User Scenario

In Fig. 5, we compare the SER performance of the proposed GMM-clustering-based approach with that of the MLD-based receiver with full CSI knowledge in a single-user scenario with QPSK modulation. As seen in this figure, when the sample size \(N\) is large, the proposed approach performs close to the optimal MLD-based one with full CSI, while at the high SNR regime, the SERs of the two approaches are almost the same. For small \(N\), there is a small difference in the performance of the two approaches, which can mainly be attributed to limited observations making the cluster boundaries of GMM sub-optimal. In addition, Fig. 5 shows that the approximation given by (18) can predict the SER of the proposed approach well.

B. Two-User NOMA

We show the performance of the proposed GMM-clustering-based approach for a two-user NOMA communication system in Fig. 6. As seen in this figure, the proposed approach performs close to the optimal MLD-based approach with full CSI. Similar to the single-user case, as the number of transmitted symbols increases, the SER performance of the proposed approach reaches that of the MLD-based approach with full CSI. An advantage of the clustering-based approach is that each user needs to send only two pilot symbols for symbol-to-bit demapping. However, in current systems, to acquire sufficiently accurate CSI at the receiver, each user needs to send a long pilot sequence that is usually more than six symbols [49]. This is inefficient when the packet size (number of transmitted symbols) is small [50]. In our proposed approach, the BS does not require a perfect estimation of the channel to eliminate the influence of channel rotation. We can find the exact quadrant of each cluster by using only two pilot symbols. Fig. 6 also shows that the theoretical predictions of (21) and (22) are reasonably accurate.

1) Impact of Difference in User Powers: When the received powers from the users are similar, the distance between the clusters associated with the weaker user increases. This can lead to four close clusters around the center each belonging to a different user. Distinguishing these clusters can be hard for the BS as they may be too close to each other or even overlap. Fig. 7 illustrates such a case. As seen in Fig. 7(a), when the difference in the power of two users is adequately large, both users can detect their signals correctly as the received signals form distinct clusters. However, as the power difference between the users decreases (Fig. 7(b) and (c)), the clusters associated with the weaker user(s) grow farther from each other. Consequently, four not-so-distinct clusters appear around the center that belong to different mappings.

In Fig. 8, we present the SER performance of the proposed approach in a two-user NOMA scenario for three values of user power difference. One can see from Fig. 8(a) that when the user power difference is sufficiently large, the BS can detect the symbols correctly even with low SNR. However, as the user power difference decreases, the SER deteriorates for both users. The SER performance of a two-user NOMA when the user power difference is relatively low is shown in Fig. 8(b) and (c). We observe that the proposed algorithm can detect the signals of both users even when the users’ received signal powers at the BS are close to each other. It is important to note that in power-domain

Fig. 5. SER comparison of the proposed GMM-clustering-based and optimal MLD-based approaches for point-to-point communication. (a) \(N = 500\), (b) \(N = 200\), and (c) \(N = 100\).
Fig. 6. SER comparison of the proposed GMM-clustering-based and optimal MLD-based approaches for two-user NOMA when $\gamma_1 - \gamma_2 = 9$ dB. (a) $N = 500$, (b) $N = 200$, and (c) $N = 100$.

Fig. 7. Constellation of received signals in a two-user NOMA scenario for three values of user power difference when $\gamma_1 = 10$ dB. (a) $\gamma_1 - \gamma_2 = 7$ dB, (b) $\gamma_1 - \gamma_2 = 5$ dB, and (c) $\gamma_1 - \gamma_2 = 3$ dB.

Fig. 8. SER performance of the proposed GMM-clustering-based and optimal MLD-based approaches for the considered two-user NOMA scenario. (a) $\gamma_1 - \gamma_2 = 7$ dB, (b) $\gamma_1 - \gamma_2 = 5$ dB, and (c) $\gamma_1 - \gamma_2 = 3$ dB.

NOMA, the power of the signals from different users should be substantially different. Otherwise, the BS will not be able to distinguish the signals. This is clear in Fig. 8(c), which shows that, when the user power difference is 3 dB, even the optimal MLD-based receiver performs poorly.

2) Comparison With Semi-Blind Estimation: We have so far compared our proposed joint estimation and detection approach with the optimal approach based on MLD with full CSI. However, in real-world scenarios, obtaining full CSI with only a limited number of symbols is infeasible. Fig. 9 shows the performance of the proposed algorithm when two symbols are used for demapping in comparison with the MLD-based approach with full CSI. MLD-based approach using two training symbols for channel estimation, and MLD-based approach using eight training symbols for channel estimation. As the results show, the performance of MLD with two symbols used for channel estimation is significantly inferior to the proposed algorithm using the same two pilot symbols. For the MLD-based approach to attain a performance close to that of our proposed algorithm, it requires at least eight pilot symbols to estimate the channel sufficiently accurately. This means, for a packet length of 100 symbols, the proposed algorithm offers at least six percent improvement in throughput. While the MLD-based joint estimate and detection approach has a low pilot-to-frame.
length in the context of long packet communication, for short packet communication, as shown in Fig. 9, the pilot-to-frame length of MLD is higher compared to our proposed algorithm at a similar SER.

Fig. 10 shows the performance of our proposed algorithm in comparison with a receiver based on the semi-blind channel estimation method proposed in [18]. We observe that, with the same number of pilot symbols, the proposed algorithm offers considerably better performance for all users. For the strongest user, the approach based on semi-blind channel estimation needs at least eight pilot symbols to perform on a par with the proposed algorithm. This means the proposed algorithm can deliver at least six percent higher throughput. Note that the semi-blind approach using ten pilot symbols may offer better SER performance for the weaker user.

3) Impact of the Convergence Threshold $\epsilon$: One of the parameters that should be considered when using the proposed algorithm is the convergence threshold $\epsilon$. This parameter has a direct impact on the time complexity of the proposed algorithm as well as its SER performance. As shown in Fig. 11, when the value of $\epsilon$ is small, not only the speed of the algorithm decreases, but also its performance degrades due to possible over-estimation. As the convergence threshold increases, both the SER performance and convergence speed improve. However, the threshold ought to be set carefully as increasing it excessively may result in under-estimation.

4) Impact of Pilot Symbol Number on Symbol-to-Bit Demapping: As we mentioned earlier, in the high SNR regime and when the received powers from the users are sufficiently different, a single pilot symbol is enough for demapping. With the QPSK modulation and using the one-symbol demapping pilot, we can identify the clusters belonging to each quadrant. Afterwards, considering the centroid of each cluster, we can calculate the phase and modulus of each channel. We demonstrate the performance of the one-symbol-based detection in Fig. 12. It is evident that, by using one symbol for demapping, the SER performance for the weaker user can approach that of the MLD-based approach with full CSI. In addition, there is less than 1 dB difference between the SER performance for the weaker user and the optimal MLD-based detection.
Fig. 13. Performance comparison of the proposed algorithm with the algorithm proposed in [48].

Fig. 14. SER performance of the proposed GMM-clustering-based approach and the MLD-based approach with full CSI for a three-user NOMA scenario when \( \gamma_1 - \gamma_2 = 9 \text{dB} \), \( \gamma_2 - \gamma_3 = 9 \text{dB} \), \( N = 500 \), and \( \epsilon = 5 \).

5) Comparison With Existing Works: The error performance of uplink NOMA has been investigated in previous research [48], [51], [52]. In Fig. 13, we compare the proposed algorithm with the approach of [48] where deep learning is utilized for multi-user detection in uplink NOMA. We consider that each user employs four pilot symbols. In our algorithm, we use a packet length of 100 symbols, while in [48], the authors consider packets of 12 symbols. As seen in Fig. 13, our algorithm performs similar to [48].

C. Higher Numbers of Users

To evaluate the performance of the proposed algorithm when the number of users increases, we consider a three-user NOMA scenario. Fig. 14 shows the SER performance of the proposed algorithm for three-user NOMA when the number of symbols is \( N = 500 \) and three pilot symbols are used for symbol-to-bit demapping. It is clear that the proposed algorithm detects the signals with good accuracy. We note that in power-domain NOMA, usually two or three users are paired. This is because the SIC at the receiver cannot successfully detect user signals, if their powers do not differ sufficiently [53]. Maintaining large power differences among a large number of paired signals is impractical.

D. Higher-Order Modulations

To demonstrate that the proposed algorithm can be used with any modulation scheme, we consider a two-user NOMA scenario where user 1 utilizes the 16-QAM modulation scheme and user 2 utilizes the QPSK modulation scheme. As before, we benchmark the SER performance of the proposed algorithm, which uses two pilot symbols, against that of the optimal MLD-based approach with full CSI. Fig. 15 shows that the proposed algorithm with only two pilot symbols performs nearly as well as the MLD-based approach with full CSI. For the MLD-based approach, the BS needs to obtain the CSI for all users using long pilot sequences. However, the proposed algorithm only needs two pilot symbols, which are used to determine the symbol-to-bit demapping.

VI. GRANT-FREE TRANSMISSION: A PRACTICAL SCENARIO

Thus far, we have assumed that the number of users is fixed and the receiver knows this number. This information is used to determine the number of clusters in our proposed algorithm. In this section, we show that the proposed algorithm can be implemented in a real-world scenario where the receiver does not have any prior knowledge except for the modulation scheme used by the transmitters and their power levels. To ensure frame synchronization, we assume that the BS periodically broadcasts beacon signals, allowing users to align their internal clocks and coordinate transmissions [54].

We slightly modify the proposed GMM-clustering-based algorithm (Algorithm 1) to cope with not knowing the number of users. The modified algorithm is summarized in Algorithm 2. In particular, we first find the average signal power \( \mathbb{E}[|y|^2] \). If it is larger than the noise power (line 2), we perform GMM clustering to assign the data into \( M \) clusters and detect the user signals. After applying SIC, we reevaluate the signal power and continue the above process until the signal power falls below the
We have not considered adaptive modulation and coding (AMC) in our system model. To use AMC, the devices need to have an estimate of the channel to choose the proper modulation order and code rate. This means that their channels need to be estimated at the BS first and then feedback to the devices, or in the case of TDMA, they can estimate their channels themselves. However, this is not realistic for mIT due to a large number of devices and the fact that they want to send short payload data only. Moreover, in our proposed approach, we assume that the devices and the BS do not know the channels in advance; therefore, it is not possible to perform AMC.

We consider a grant-free scenario where the number of active users is not known to the BS. Each user randomly selects a power level from a pool of possible values $\mathbb{P}_p = \{P_1, P_2, \ldots, P_5\}$, and there is no cooperation between the users in this regard.

We assume each user performs power control such that its average received power at the BS equals the selected power level. The received signal at the BS at time instant $i$ is then given by

$$y = \sum_{i=1}^{K} \sqrt{P_i} g_i x_i + \epsilon,$$

where $p_i \in \mathbb{P}_p$, $g_i$ represents small-scale fading, $w \sim CN(0, 1)$, and $E[|x_i|^2] = 1$. The BS does not have any prior knowledge of $K$, $g_i$, or $p_i$, but knows $\mathbb{P}_p$, $N_0$, and the modulation scheme utilized by the users.

We first ignore the small-scale fading, i.e., we assume $g_i = 1$. We also consider five power levels calculated as $P_i = P_1 + (i - 1) \times 7$ dB for $i = 1, \ldots, 5$. The number of active users in each time frame (with the duration of one packet of length $N$ symbols) is uniformly distributed between one and three. At each timeslot, a random number of users start transmitting their signals. Fig. 16 shows the SER performance of Algorithm 2 when we disregard the small-scale fading and users apply power control. As can be seen in Fig. 16, the performance of our proposed model is matched with that of MLD.

We then consider a real-world scenario under the Rayleigh fading scenario ($g_i \sim CN(0, 1)$), where each user randomly selects one of the power levels as its transmit power and starts transmitting its packet. We only consider power control as a means for mitigating the effects of large-scale fading. To simulate a practical real-world scenario, we further consider a 5 dB difference in power between levels. Fig. 17 shows the SER of grant-free transmission for a SIC receiver based on Algorithm 2 and its comparison with the MLD with full CSI. As mentioned above, we consider five power levels, and each user randomly selects one of the power levels for communication with the BS. At the receiver, we apply Algorithm 2 and average the users’ SERs based on their power levels. It is clear that the performance of the proposed algorithm using only three pilot symbols is close to that of MLD with full CSI.

Fig. 18 shows the throughput performance of the grant-free transmission with a SIC receiver based on Algorithm 2 and its comparison with the MLD-based receiver with full CSI. At the receiver, we apply Algorithm 2 and calculate the average throughput based on the number of users. The average throughput is defined as the ratio of the number of symbols correctly detected at the receiver and the total number of transmitted symbols. In particular, we assume that a random number of users start transmitting their packets to the BS. BS applies Algorithm 2, detects each user’s signals, and calculates the SER of each user.

Let us assume the SER of user $i$ as $e_i$, representing the ratio of
erroneous symbols to the packet length. The system’s average SER is then expressed as $\sum^{N_u}_{i=1} e_i N_u \epsilon_i$ where $N_u$ denotes the total number of users. Following this, we calculate the throughput of the system using $1 - \sum^{N_u}_{i=1} e_i N_u \epsilon_i$ [61]. In order to obtain the outcomes depicted in Fig. 18, we performed a series of 1000 simulation iterations and calculated the average results. As seen in Fig. 18, the performance of the proposed algorithm with only three pilot symbols is close to that of the MLD-based approach with full CSI that has the knowledge of a number of users. For a fair comparison, we also simulated the case where MLD is imperfect, BS lacks information regarding the number of users, and the number of pilots is the same as our proposed GMM-based model. As can be seen, our proposed algorithm has significantly superior performance for the identical simulation setup.

VII. CONCLUSION

In this article, we proposed to employ a semi-supervised machine learning algorithm, i.e., GMM clustering, to cluster the signals at the receiver for joint channel estimation and signal detection in grant-free NOMA. We applied an SIC strategy to detect the signals of each user. We compared the performance of our proposed GMM-clustering-based algorithm with that of the optimal detection method based on MLD, which requires full CSI at the receiver. We showed that using a single pilot symbol for each user, the proposed algorithm can reach the performance of the MLD-based approach with full CSI. We took one step further and used a single pilot symbol to estimate the channel of all users. The resulting performance was still close to that of the MLD with full CSI. To make a fairer comparison, we also compared the performance of our proposed algorithm with that of the MLD with imperfect channel estimation relying on a finite number of symbols. We observed that, to attain a similar performance, the MLD-based approach needs at least eight pilot symbols for channel estimation to perform as well as our proposed algorithm using only two pilot symbols. Furthermore, to reduce the computational burden at the BS, we proposed a theoretical model to calculate the probability of error based on the maximum error of the EM algorithm utilized for GMM clustering. Our simulation results showed that the proposed model predicts the SER of the proposed algorithm well. The simulation results also demonstrated that, when the number of transmitted symbols is moderate or large, the SER performance of the proposed algorithm is on a par with that of the optimal MLD. Finally, since the accuracy of the GMM clustering depends on the sample size, we showed that there exists a tradeoff between the accuracy of the proposed algorithm and the communication block length.

APPENDIX

PROOF OF THEOREM 1

The proof is based on the findings of [32], [33], [62]. Without loss of generality, we focus on the update rule for one of the centers. We start by writing the update rule for the mean as

$$
\mu_i^+ - \mu_i^* = \frac{\mathbb{E}[\gamma_1(X, \mu_1) (X - \mu_i^+)]}{\mathbb{E}[\gamma_1(X, \mu_1)]}.
$$

Since the vector of the true centers $\mu^*$ is fixed, we have

$$
\mathbb{E}[\gamma_1(X, \mu_1^*) (X - \mu_i^+)] = 0.
$$

Hence, we can write
\[
\mu^*_1 - \mu^*_1 = \frac{\mathbb{E} [(\gamma_1(X, \mu) - \gamma_1(X, \mu^*)) (X - \mu^*_1)]}{\mathbb{E} [\gamma_1(X, \mu)]}.
\] (28)

We find an upper bound on the norm of the expectation in the numerator. Therefore, we define
\[
\mu^t := \mu^* + t(\mu - \mu^*)
\]
and
\[
g_X(t) := \gamma_1(X, \mu^t).
\]
Subsequently, we have
\[
\gamma_1(X, \mu) - \gamma_1(X, \mu^*) = \int_0^1 g_X(t) dt
\]
\[
= \int_0^1 \nabla \gamma_1(x, \mu^1_t)^T (\mu^*_1 - \mu^*_1) dt.
\] (29)

Computing the integration and applying the expectation, the upper bound can be written as
\[
\|\mathbb{E} [(\gamma_1(X, \mu) - \gamma_1(X, \mu^*)) (X - \mu^*_1)]\|_2
\]
\[
\leq V_i \|\mu_1 - \mu^*_1\|_2 + \sum_{i \neq 1} V_i \|\mu_i - \mu^*_i\|_2
\]
\[
\leq M (\max_i V_i) (\max_i \|\mu_i - \mu^*_i\|_2)
\] (30)

where
\[
V_i = \sup_{t \in [0, 1]} \|\mathbb{E} [(\gamma_1(X; \mu^t) - 1) - \gamma_1(X; \mu^t))] (X - \mu^*_1)\|_2
\]
\[
\times (X - \mu^*_1)^T \|_{op}
\]
\[
V_i = \sup_{t \in [0, 1]} \|\mathbb{E} [(\gamma_1(X; \mu^t) - \gamma_1(X; \mu^t)) (X - \mu^*_1) - \gamma_1(X; \mu^t)^T \|_{op}
\] (31)

Considering \(Z\) as the label of \(X\), one can write
\[
\mathbb{E} [\gamma_1(X; \mu^t)] = \mathbb{E} [\mathbb{P}_M (Z = 1 | X)] = \omega_1 > \kappa.
\] (32)

When \(\mu\) is in the vicinity of \(\mu^*\), we have \(\mathbb{E} [\gamma_1(X; \mu) | \mu \approx \mathbb{E} [\gamma_1(X; \mu^*)] > \kappa\). According to [32, Lemma 5.2], we know that, as long as \(R_{\min} \geq 30 \min \{M, d\}\) and
\[
a \geq \frac{1}{2} R_{\min} - \min \{M, d\}\) \(0.5
\]
\[
\times \max \left\{ 4\sqrt{2} \left[ \log \left( \frac{R_{\min}}{4} \right) \right]^{\frac{1}{2}} \right\} + \frac{\sqrt{\frac{3}{2}} \log \left( \frac{R_{\min}}{4} \right)}{\kappa}
\] (33)

for any \(\mu_i \in B(\mu^*_i) \) \(i \in [M]\), we have \(\mathbb{E} [\gamma_1(X; \mu)] \geq \frac{1}{2} \kappa \) \(i \in [M]\).

Using the above result and the upper bound (30), we have
\[
\|\mu^*_1 - \mu^*_1\|_2 = \|\mathbb{E} [(\gamma_1(X, \mu) - \gamma_1(X, \mu^*)) (X - \mu^*_1)]\|_2
\]
\[
\leq 4 \frac{M}{3 \kappa} (\max_i V_i) (\max_i \|\mu_i - \mu^*_i\|_2).
\] (36)

Defining the event \(\mathcal{E}_{1,i}\) as
\[
\sup_{\mu \in \mathcal{U}} \left\| \frac{1}{n} \sum_{j=1}^n \gamma_i(X_j; \mu) (X_j - \mu^*_i) - \mathbb{E} [\gamma_i(X; \mu) (X - \mu^*_i)] \right\|_2
\]
\[
\leq 1.5 R_{\max} \left( \frac{C_3 M d \log n}{n} \right)^{0.5}
\] (37)

and the event \(\mathcal{E}_{2,i}\) as
\[
\sup_{\mu \in \mathcal{U}} \left| \frac{1}{n} \sum_{j=1}^n \gamma_i(X_j; \mu) - \mathbb{E} [\gamma_i(X; \mu)] \right| \leq \left( \frac{C_2 M d \log n}{n} \right)^{0.5}
\] (38)

where
\[
\mathcal{U} = \prod_{i=1}^M B(\mu^*, R_{\max})
\]
\[
\mathcal{C}_2 = C_2 \log \left( M (2 \sqrt{2} R_{\max} + \sqrt{d}) \right)
\]
\[
\mathcal{C}_3 = C_3 \log \left( M \left( 6 R_{\max}^2 + \sqrt{d} \right) \right),
\]
and \(C_2\) and \(C_3\) are universal constants.

In view of [32, Lemmas 5.3 and 5.4], the event \(\{ \cap_{i \in [M]} \mathcal{E}_{1,i} \} \cap \{ \cap_{i \in [M]} \mathcal{E}_{2,i} \} \) for all \(i \in [M]\) occurs with the probability at least \(1 - \frac{2}{\kappa} \). Due to the sample size condition (15), we have
\[
R_{\max} \left( \frac{C_3 M d \log n}{n} \right)^{0.5} \leq \frac{\kappa}{\max_{i \in [M]} \|\mu_i^0 - \mu^*_i\|_2}
\] (39)
\[
\left( \frac{C_2 M d \log n}{n} \right)^{0.5} \leq \frac{\kappa}{12}.
\] (40)

Using the definition of the event \(\mathcal{E}_{2,i}\) for all \(i \in [M]\), the second inequality (40) can be written as
\[
\sup_{\mu \in \mathcal{U}} \left| \frac{1}{n} \sum_{j=1}^n \gamma_i(X_j; \mu) - \mathbb{E} [\gamma_i(X; \mu)] \right| \leq \frac{\kappa}{12}.
\] (41)

Taking \(\mu^{(0)}\) as the initial value of the mean, we can write
\[
\|\mu^{(1)}_i - \mu^*_i\|_2 = \frac{1}{n} \sum_{j=1}^n \gamma_i(X_j; \mu^{(0)})(X_j - \mu^*_i)\|_2
\]
\[
\leq \frac{1}{n} \sum_{j=1}^n \gamma_i(X_j; \mu^{(0)}) \|X - \mu^*_i\|_2 + R_{\max} \left( \frac{C_3 M d \log n}{n} \right)^{0.5}
\]
\[
\leq \frac{\|\mathbb{E} [\gamma_i(X; \mu^{(0)})] - X - \mu^*_i\|_2 + R_{\max} \left( \frac{C_3 M d \log n}{n} \right)^{0.5}}{\|\gamma_i(X; \mu^{(0)})\|_2}
\] (43)
\[
\leq \frac{\|\mathbb{E} [\gamma_i(X; \mu^{(0)})] - X - \mu^*_i\|_2 + R_{\max} \left( \frac{C_3 M d \log n}{n} \right)^{0.5}}{\|\gamma_i(X; \mu^{(0)})\|_2}
\]
\[
\leq \frac{1}{\min_{i \in [M]} \|\mu^{(0)}_i - \mu^*_i\|_2} + 3 R_{\max} \left( \frac{C_3 M d \log n}{n} \right)^{0.5}.
\] (45)
Using the above inequality and the first sample size condition, we have
\[
\|\mu_i^{(t)} - \mu_i^*\|_2 \leq \frac{1}{2} \max_{i \in [M]} \|\mu_i^{(0)} - \mu_i^*\|_2 + \frac{3R_{\max} \left( \frac{C_i M_\delta \text{log}(n)}{n} \right)^{0.5}}{2\kappa} \leq \max_{i \in [M]} \|\mu_i^{(0)} - \mu_i^*\|_2.
\]
By applying (42)–(45) over \(t\) iterations, we have
\[
\begin{align*}
\max_{i \in [M]} \|\mu_i^{(t)} - \mu_i^*\|_2 & \leq \frac{1}{2} \max_{i \in [M]} \|\mu_i^{(0)} - \mu_i^*\|_2 \left( 1 + \frac{1}{2} + \cdots + \frac{1}{2^{t-1}} \right) \\
& \leq \frac{1}{2} \max_{i \in [M]} \|\mu_i^{(0)} - \mu_i^*\|_2 + \frac{3R_{\max} \left( \frac{C_i M_\delta \text{log}(n)}{n} \right)^{0.5}}{2\kappa}.
\end{align*}
\]

REFERENCES

[1] M. Series, “Minimum requirements related to technical performance for IMT-2020 radio interface (n),” ITU Geneva, Switzerland, Report ITU-R, M.2410-0, 2017.
[2] M. Framingham, “Growth in connected IoT device,” in Proc. Int. Data Corporation, 2019, pp. 1–9.
[3] M. Vaezi et al., “Cellular, wide-area, and non-terrestrial IoT: A survey on 5G advances and the road toward 6G,” IEEE Commun. Surv. Tut., vol. 24, no. 2, pp. 1117–1174, Secondquarter 2022.
[4] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, “Grant-free non-orthogonal multiple access for IoT: A survey,” IEEE Commun. Surv. Tut., vol. 22, no. 3, pp. 1805–1838, Thirdquarter 2020.
[5] M. Alikhabas, M. Toka, S. Gökdal, G. K. Kurt, and O. Kucur, “A tutorial on non-orthogonal multiple access for 5G and beyond,” Wireless Commun. Mobile Comput., vol. 2018, pp. 1–24, 2018.
[6] Z. Yin, M. Jia, W. Wang, N. Cheng, F. Lyu, and X. Shen, “Max-min secrecy rate for noma-based UAV-assisted communications with protected zone,” in Proc. IEEE Glob. Commun. Conf., 2019, pp. 1–6.
[7] Z. Yin et al., “Secrecy rate analysis of satellite communications with frequency domain NOMA,” IEEE Trans. Veh. Technol., vol. 68, no. 12, pp. 11847–11858, Dec. 2019.
[8] R. Abbas, T. Huang, B. Shahab, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, “Grant-free non-orthogonal multiple access: A key enabler for 6G-IoT,” 2020, arXiv:2003.10257.
[9] N. H. Mahmood, H. Alves, O. A. López, M. Shehab, D. P. M. Osorio, and M. Latva-Aho, “Six key features of machine type communication in 6G,” in Proc. 2nd 6G Wireless Summit, 2020, pp. 1–5.
[10] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, “The roadmap to 6G: AI empowered wireless networks,” IEEE Commun. Lett., vol. 21, no. 3, pp. 640–643, Mar. 2017.
[11] 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.
[12] 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, Jun. 2018.
[13] A. Salari, M. Shirvanimoghaddam, M. B. Shahab, R. Arablouei, and S. Johnson, “Clustering-based joint channel estimation and signal detection for grant-free NOMA,” in Proc. IEEE Globecom Workshops, 2020, pp. 1–6.
[14] A. Salari, M. Shirvanimoghaddam, M. B. Shahab, Y. Li, and S. Johnson, “NOMA joint channel estimation and signal detection using rotational invariant codes and GM-based clustering,” IEEE Commun. Lett., vol. 26, no. 10, pp. 2485–2489, Oct. 2022.
[15] R. J. Rossi, Mathematical Statistics: An Introduction to Likelihood Based Inference. Hoboken, NJ, USA: Wiley, 2018.
[16] K. P. Murphy, Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press, 2012.
[17] C. M. Bishop, Pattern recognition and machine learning (information science and statistics), 2007.
[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” J. Roy. Stat. Soc.: Ser. B (Methodol.), vol. 39, no. 1, pp. 1–22, 1977.
[19] X.-L. Meng and D. B. Rubin, “Maximum likelihood estimation via the ECM algorithm: A general framework,” Biometrika, vol. 80, no. 2, pp. 267–278, 1993.
[20] J. Xu, D. J. Hsu, and A. Maleki, “Global analysis of expectation maximization for mixtures of two Gaussians,” in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 2684–2692.
[21] R. Zhao et al., “Statistical convergence of the EM algorithm on Gaussian mixture models,” Electron. J. Statist., vol. 14, no. 1, pp. 632–660, 2020.
[22] B. Yan, M. Yin, and P. Sarkar, “Convergence of gradient EM on multi-component mixture of Gaussians,” in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 6959–6969.
[23] H. Ashiani, S. Ben-David, N. Harvey, C. Liaw, A. Mehrabian, and Y. Plan, “Nearly tight sample complexity bounds for learning mixtures of Gaussians via sample compression schemes,” in Proc. Adv. Neural Inf. Process. Syst., 2018, pp. 3416–3425.
[24] J. Kwon and C. Caramanis, “The EM algorithm gives sample-optimality for learning mixtures of well-separated Gaussians,” in Proc. Conf. Learn. Theory, 2020, pp. 2425–2487.
[25] B. Yu, Y. Cai, and D. Wu, “Joint access control and resource allocation for short-packet-based mMTC in status update systems,” IEEE J. Sel. Areas Commun., vol. 39, no. 3, pp. 851–865, Mar. 2021.
[26] S. Lv, X. Xu, S. Han, X. Tao, and P. Zhang, “Energy-efficient secure short-packet transmission in NOMA-assisted mMTC networks with relaying,” IEEE Trans. Veh. Technol., vol. 71, no. 2, pp. 1699–1712, Feb. 2022.
[27] K. F. Hasan, C. Wang, Y. Feng, and Y.-C. Tian, “Time synchronization in vehicular ad-hoc networks: A survey on theory and practice,” Veh. Commun., vol. 14, pp. 39–51, 2018.
[28] R. Singh, B. C. Pal, and R. A. Jabr, “Statistical representation of distribution system loads using Gaussian mixture model,” IEEE Trans. Power Syst., vol. 25, no. 1, pp. 29–37, Feb. 2010.
