JOINT ESTIMATION OF CLUSTERED USER ACTIVITY AND CORRELATED CHANNELS WITH UNKNOWN COVARIANCE IN MMTC

Hamza Djelouat, Markus Leinonen, and Markku Juntti

Centre for Wireless Communications – Radio Technologies, FI-90014, University of Oulu, Finland

ABSTRACT

This paper considers joint user identification and channel estimation (JUICE) in grant-free access with a *clustered* user activity pattern. In particular, we address the JUICE in massive machine-type communications (mMTC) network under correlated Rayleigh fading channels with unknown channel covariance matrices. We formulate the JUICE problem as a maximum *a posteriori* probability (MAP) problem with properly chosen priors to incorporate the partial knowledge of the UEs' clustered activity and the unknown covariance matrices. We derive a computationally-efficient algorithm based on alternating direction method of multipliers (ADMM) to solve the MAP problem iteratively via a sequence of closed-form updates. Numerical results highlight the significant improvements brought by the proposed approach in terms of channel estimation and activity detection performances for clustered user activity patterns.

Index Terms— mMTC, user identification, channel estimation, clustered activity, spatially correlated channels.

1. INTRODUCTION

Sparse signal recovery techniques have widely been applied in the design of massive machine-type communications (mMTC) solutions with grant-free access protocols. A major challenge in grant-free access is the joint user identification and channel estimation (JUICE). Subsequently, owing to the sporadic nature of activity pattern of the mMTC devices, namely user equipments (UEs), the JUICE has been widely addressed as a sparse recovery problem and solved using various algorithms such as approximate message passing (AMP), sparse Bayesian learning (SBL), and mixed-norm minimization.

The vast majority of the works address the JUICE in an uncorrelated channel model [1, 2, 3, 4, 5, 6]. However, this assumption is not practical as the multiple-input multiple-output (MIMO) channels are usually spatially correlated [7]. Recently, few works have addressed the JUICE in the correlated MIMO channels, thereby better reflecting the reality, using AMP in [8, 9] and using mixed-norm minimization in [10, 11, 12]. Nonetheless, these works assume that channel distribution information (CDI) for all the UEs is available at the base station (BS), which can be challenging to acquire in realistic environments. Furthermore, the prior work on the JUICE considers mMTC networks with a random UE activity pattern. This could model, e.g., a scenario where UEs monitor independent random processes and thus activate randomly based on certain application criteria.

This paper makes the following two distinctions from the prior works: First, we address the JUICE in spatially correlated MIMO channels with *no prior* knowledge of the exact CDI. Second, we

consider an mMTC network where the UEs activity has a specific structure, namely, a *clustered* form. For instance, this models a network where the UEs form clusters based on their geographical locations and each cluster is associated with a monitoring task. Here, an event could trigger a small subset of UEs belonging to a cluster to activate concurrently, leading to clustered UE activity system-wise.

This paper adopts a Bayesian approach and formulates the JUICE as a maximum *a posteriori* probability (MAP) estimation problem that takes into consideration the different specific structures of the system model. In particular, to accommodate the clustered activity of the UEs, we propose a log-sum based prior function composed of two different components to capture both the overall sparsity and the clustered sparsity of the signal model. Furthermore, in order to estimate the CDI as well, we utilize a prior function that follows a Wishart distribution. Subsequently, we derive a computationally-efficient iterative solution based on alternating direction method of multipliers (ADMM) that solves an approximate version of the MAP problem as a sequence of closed-form updates.

2. SYSTEM MODEL

We consider a single-cell uplink network consisting of N UEs served by a single BS equipped with a uniform linear array (ULA) of M antennas. The UEs are geographically distributed so that they form C *clusters*, where each cluster contains L UEs[1], such that $N = LC$. A cluster containing a subset of UE indices is denoted by $\mathcal{C}_l \subseteq \{1, 2, \ldots, N\}$.

The propagation channel between the ith UE and the BS, denoted by $\mathbf{h}_i$, follows a local scattering model [13]. Thus, each $\mathbf{h}_i$ is modelled as a zero-mean complex Gaussian random variable, i.e., $\mathbf{h}_i \sim \mathcal{CN}(0, \boldsymbol{\Sigma}_i^{-1})$, where $\boldsymbol{\Sigma}_i \in \mathbb{C}^{M \times M}$ denotes the channel precision matrix, i.e., the inverse covariance matrix $\boldsymbol{\Sigma}_i^{-1} = \mathbb{E}[\mathbf{h}_i \mathbf{h}_i^{\mathsf{H}}]$. We consider UEs with low mobility, which is justified in mMTC, e.g., in applications like industrial monitoring. Hence, we adopt the common assumption that the channels are wide-sense stationary. Thus, the set of channel precision matrices $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_i\}_{i=1}^N$ vary in a slower timescale compared to the channel realizations [7].

Since the UE activation is sporadic in mMTC, only a small number of the N UEs are active at each coherence time T_{c}. To this end, the UE activity indicator vector $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_N]^{\mathsf{T}}$ is given as

$$\gamma_i = \begin{cases} 1, & i\text{th UE is active} \\ 0, & \text{otherwise,} \end{cases} \qquad \forall i = 1, \ldots, N. \qquad (1)$$

Contrary to the majority of the literature on mMTC that consider random UE activation, we consider that the UEs activate in a specific *clustered* manner. More precisely, the active UEs belong to a small (random) number of active clusters, where an active cluster refers

---

[1]For the simplicity of presentation, yet without loss of generality, we assume that all clusters contain the same number of UEs.

to any cluster with at least one active UE, while containing at most $L_c \leq L$ active UEs. Such cluster sparsity imposes a special structure on $\boldsymbol{\gamma} = [\underbrace{\gamma_1, \ldots, \gamma_L}_{\text{Cluster 1}}, \underbrace{\gamma_{L+1}, \ldots, \gamma_{2L}}_{\text{Cluster 2}}, \ldots, \underbrace{\gamma_{N-L+1}, \ldots, \gamma_N}_{\text{Cluster } C}]$ where the elements belonging to a given cluster ($\gamma_i, i \in \mathcal{C}_l$) are assumed to be correlated. The correlation in the UE activity is imposed by the priors presented in Section 3.3.

For data transmission, each UE $i \in \mathcal{N}$ is assigned a unique unit-norm pilot sequence $\boldsymbol{\phi}_i \in \mathbb{C}^{\tau_P}$, as well as a transmit power p_i that is inversely proportional to its average channel gain [4, 14]. Accordingly, the received pilot signal $\mathbf{Y} \in \mathbb{C}^{\tau_P \times M}$ is given by

$$\mathbf{Y} = \sum_{i=1}^{N} \gamma_i \sqrt{p_i} \boldsymbol{\phi}_i \mathbf{h}_i^{\mathsf{T}} + \mathbf{W} = \boldsymbol{\Phi} \mathbf{X}^{\mathsf{T}} + \mathbf{W}, \qquad (2)$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{C}^{M \times N}$ denotes the effective channel matrix with $\mathbf{x}_i = \gamma_i \sqrt{p_i} \mathbf{h}_i$, $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N] \in \mathbb{C}^{\tau_P \times N}$ is the pilot book, and $\mathbf{W} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_M) \in \mathbb{C}^{\tau_P \times M}$ represents additive white Gaussian noise.

3. JUICE WITH CLUSTERED ACTIVITY AND PARTIALLY KNOWN CDI

The signal model in (2) represents a linear measurement where the (unknown) effective channel matrix $\mathbf{X}^{\mathsf{T}}$ exhibits row sparsity. Thus, detecting the active UEs and estimating their channel gives rise to a sparse recovery problem from a multiple measurement vector (MMV) setup, which can be solved via several CS algorithms. For instance, the Bayesian AMP algorithm [15] can be utilized if the statistics of a sparse signal are known a priori. However, when such information is not available, mixed-norm minimization [16] using deterministic sparsity regularization is often utilized. Clearly, the additional prior information in AMP leads to superior performance in general. However, attaining prior CDI is a nontrivial task and requires network resources to perform a training phase.

In this paper, we aim to tackle the aforementioned issue related to the acquisition of CDI by adopting a Bayesian framework which aims at not only detecting the active UEs and estimate their channels, but also estimating their covariance matrices. More precisely, by formulating the JUICE as an MAP estimation problem and using the adequate prior functions, we develop a computationally-efficient algorithm that solves a relaxed version of the MAP problem iteratively via a set of closed-form update rules, as we show next.

The MAP estimates $\{\hat{\mathbf{X}}, \hat{\gamma}, \hat{\boldsymbol{\Sigma}}\}$ with respect to the posterior density given the measurement matrix $\mathbf{Y}$ is given by

$$\begin{aligned} \{\hat{\mathbf{X}}, \hat{\gamma}, \hat{\boldsymbol{\Sigma}}\} &= \max_{\mathbf{X}, \gamma, \boldsymbol{\Sigma}} p(\mathbf{X}, \gamma, \boldsymbol{\Sigma} | \mathbf{Y}) \\ &= \max_{\mathbf{X}, \gamma, \boldsymbol{\Sigma}} \frac{p(\gamma, \boldsymbol{\Sigma}) p(\mathbf{X} | \gamma, \boldsymbol{\Sigma}) p(\mathbf{Y} | \mathbf{X}, \gamma, \boldsymbol{\Sigma})}{p(\mathbf{Y})} \\ &\overset{(a)}{=} \max_{\mathbf{X}, \gamma, \boldsymbol{\Sigma}} p(\gamma, \boldsymbol{\Sigma}) p(\mathbf{X} | \gamma, \boldsymbol{\Sigma}) p(\mathbf{Y} | \mathbf{X}) \\ &= \min_{\mathbf{X}, \gamma, \boldsymbol{\Sigma}} -\log p(\mathbf{Y} | \mathbf{X}) - \log p(\mathbf{X} | \gamma, \boldsymbol{\Sigma}) - \log p(\gamma, \boldsymbol{\Sigma}) \\ &\overset{(b)}{=} \min_{\mathbf{X}, \gamma, \boldsymbol{\Sigma}} \frac{1}{\sigma^2} \|\mathbf{Y} - \boldsymbol{\Phi} \mathbf{X}^{\mathsf{T}}\|_{\mathrm{F}}^2 - \log p(\mathbf{X} | \gamma, \boldsymbol{\Sigma}) \\ &\qquad - \log p(\boldsymbol{\Sigma}) - \log p(\gamma) \end{aligned}$$

$$\tag{3}$$

where (a) follows from the Markov chain $\{\gamma, \boldsymbol{\Sigma}\} \to \mathbf{X} \to \mathbf{Y}$ and because $p(\mathbf{Y})$ does not affect the maximization and (b) follows from the additive Gaussian noise model in (2) and independence of UE activity and channels. Next, we elaborate the definition of the conditional PDF $p(\mathbf{X}|\gamma)$ and the choice of the priors $p(\boldsymbol{\Sigma})$ and $p(\gamma)$.

3.1. Conditional PDF $p(\mathbf{X}|\gamma, \boldsymbol{\Sigma})$

Since the user activity is controlled by γ, the conditional PDF $p(\mathbf{x}_i | \gamma_i, \boldsymbol{\Sigma}_i)$ is defined as follows:

$$p(\mathbf{x}_i | \gamma_i, \boldsymbol{\Sigma}_i) = \mathcal{CN}(\mathbf{x}_i; \mathbf{0}, p_i \boldsymbol{\Sigma}_i^{-1})^{\gamma_i} \mathbb{I}(\mathbf{x}_i = \mathbf{0})^{1-\gamma_i}, \qquad (4)$$

where $\mathbb{I}(a)$ is an indicator function that takes the value 1 if $a \neq 0$, and 0 otherwise. $p(\mathbf{x}_i | \gamma_i, \boldsymbol{\Sigma}_i)$ implies that conditioned on $\gamma_i = 0$, $\mathbf{x}_i$ is equal to 0 with probability one, and conditioned on $\gamma_i = 1$, $\mathbf{x}_i$ follows a complex Gaussian distribution with zero mean and precision matrix $\boldsymbol{\Sigma}_i^{-1}$. Furthermore, since $\mathbf{x}_i$ depends only on γ_i and $\boldsymbol{\Sigma}_i$, $-\log p(\mathbf{X}|\gamma, \boldsymbol{\Sigma}) = \sum_{i=1}^{N} -\log p(\mathbf{x}_i | \gamma_i, \boldsymbol{\Sigma}_i)$ is given as

$$-\log p(\mathbf{X}|\gamma, \boldsymbol{\Sigma}) \propto \sum_{i=1}^{N} -\gamma_i p_i^M \log |\boldsymbol{\Sigma}_i| + \frac{\gamma_i}{p_i} \mathbf{x}_i^{\mathsf{H}} \boldsymbol{\Sigma}_i \mathbf{x}_i. \qquad (5)$$

3.2. Prior on Precision Matrix

A common and physically grounded prior for an unknown precision matrix $\boldsymbol{\Sigma}_i$ of the Gaussian random variable $\mathbf{x}_i$ is given by the Wishart distribution [17], defined as

$$p(\boldsymbol{\Sigma}_i) \sim \mathcal{W}(\boldsymbol{\Sigma}_i | \mathbf{B}_i, v) = f(\mathbf{B}_i, v) |\boldsymbol{\Sigma}_i|^d \exp\left(-\mathrm{Tr}(\mathbf{B}_i^{-1} \boldsymbol{\Sigma}_i)\right) \quad (6)$$

where $f(\mathbf{B}_i, v)$ is a normalization constant given by [17, eq (B.79)], $d = v - M + 1 > 0$ where v controls the degrees of freedom of the distribution, and $\mathbf{B}_i \in \mathbb{C}^{M \times M}$ is a symmetric, positive definite matrix that represents the prior guess for the precision matrix $\boldsymbol{\Sigma}_i$. Subsequently, we write

$$-\log p(\boldsymbol{\Sigma}) = \sum_{i=1}^{N} -\log p(\boldsymbol{\Sigma}_i) \propto \sum_{i=1}^{N} -d \log |\boldsymbol{\Sigma}_i| + \mathrm{Tr}(\mathbf{B}_i^{-1} \boldsymbol{\Sigma}_i), \quad (7)$$

where $-\log(f(\mathbf{B}, v))$ was dropped as it does not depend on $\boldsymbol{\Sigma}_i$.

3.3. Sparsity-promoting Prior $p(\gamma)$

For a sparse signal recovery problem, utilizing prior functions that promote sparsity while incorporating the possible specific structures of the sparse signal is the key to achieve accurate solutions. We discuss two different prior functions that capture the different structural features of our underlying sparse signal below.

1) Separable prior: Since the UEs have a sparse activity pattern, the optimal sparsity-inducing prior is the ℓ_0-norm penalty on γ, i.e., $\sum_{i=1}^{N} \mathbb{I}(\gamma_i)$. However, the ℓ_0-norm penalty is intractable for large N. Thus, many surrogate functions have been proposed to relax it, for instance, *log-sum* penalty $\sum_{i=1}^{N} \log(\gamma_i + \epsilon)$, which is used herein. This resembles most closely the ℓ_0-norm penalty when $\epsilon_0 \to 0$. Subsequently, we define a sparsity prior function as

$$J_s(\gamma) \propto \sum_{i=1}^{N} \log(\gamma_i + \epsilon_0). \qquad (8)$$

2) Cluster-sparsity-promoting prior: Although the prior (8) is an appropriate choice as it: 1) promotes sparsity, 2) is separable across the UEs, it ignores the considered clustered structure of the UEs' activity pattern. To account for this correlated activity, we propose a cluster-sparsity-promoting prior that captures the relations between the UE activity indicators belonging to the same cluster, i.e., $\gamma_i, i \in \mathcal{C}_l$. More precisely, we propose the following prior function:

$$J_c(\gamma) \propto \sum_{l=1}^{C} \log\left(\sum_{i \in \mathcal{C}_l} \gamma_i + \epsilon_0\right). \qquad (9)$$

While $J_\text{s}(\boldsymbol{\gamma})$ is blind to the cluster sparsity structure, $J_\text{c}(\boldsymbol{\gamma})$ promotes quite stringently solutions that have clustered sparsity as it has the tendency to enforce all UEs within each cluster to be detected active even if only one UE is active, being thereby susceptible to high false alarm error rate. Therefore, $J_\text{c}(\boldsymbol{\gamma})$ would face robustness issues in the instances where the UEs activity pattern does not exhibit a clustered structure.

4. PROPOSED SOLUTION VIA ADMM

For deriving the proposed solution, we make two technical choices:

1) Inspired by the argument raised in [18], aiming to estimate all N precision matrices $\boldsymbol{\Sigma}_i$ with the available data at the BS may lead to overfitting. Thus, we restrict the estimation of the N precision matrices $\boldsymbol{\Sigma}_i$ to L precision matrices $\boldsymbol{\Sigma}_l$. More precisely, we assume that all propagation channels of the UEs within a single cluster share the same CDI, i.e., $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_l^{-1})$, $\forall i \in \mathcal{C}_l$, $\forall l = 1, \ldots, C$.

2) The binary nature of $\boldsymbol{\gamma}$ renders the MAP estimation problem intractable for large N. To overcome this challenge, we note that finding the index set $\{i \mid \gamma_i \neq 0, \ i \in \mathcal{N}\}$ is equivalent to finding the index set $\{i \mid \|\mathbf{x}_i\| > 0, \ i \in \mathcal{N}\}$. Thus, we can eliminate the variable $\boldsymbol{\gamma}$ from the optimization problem by approximating each γ_i by $\|\mathbf{x}_i\|$ and by relaxing $p(\boldsymbol{\gamma})$ by an equivalent prior function $p(\mathbf{X})$ that depends on $\|\mathbf{x}_i\|$, $\forall i \in \mathcal{N}$, as we will show in the next section.

By using the aforementioned arguments and substituting (5) and (7) into (3), the MAP estimation problem can be rewritten as

$$\{\hat{\mathbf{X}}, \hat{\boldsymbol{\Sigma}}\} = \min_{\mathbf{X}, \boldsymbol{\Sigma}} \frac{1}{2}\|\mathbf{Y} - \boldsymbol{\Phi}\mathbf{X}^\text{T}\|_\text{F}^2 - \beta_1 \log p(\mathbf{X}) + \beta_2 \sum_{l=1}^{C} \sum_{i \in \mathcal{C}_l} \frac{\mathbf{x}_i^\text{H} \boldsymbol{\Sigma}_l \mathbf{x}_i}{p_i}$$
$$- \beta_2 \sum_{l=1}^{C} \log|\boldsymbol{\Sigma}_l| \sum_{i \in \mathcal{C}_l} p_i^M \|\mathbf{x}_i\| - \beta_3 L \sum_{l=1}^{C} \left(d \log|\boldsymbol{\Sigma}_l| + \text{tr}(\mathbf{B}_l^{-1}\boldsymbol{\Sigma}_l)\right),$$
(10)

where regularization weights β_1, β_2, and β_3 control the emphasis on the priors with respect to the measurement fidelity term.

In the following, we propose an iterative solution by alternating $p(\mathbf{X})$ over $J_\text{c}(\cdot)$ and $J_\text{s}(\cdot)$. The central idea is to develop an iterative two-level algorithm, whose outer loop aims at detecting the active clusters, and the inner loop aims at detecting the active UEs in each of the estimated active cluster. More precisely, in the outer loop, the algorithm enforces the detection of active clusters via the cluster-sparsity-promoting prior J_c in (9). Subsequently, the algorithm runs an inner loop over the just-estimated active clusters to detect the individual active UEs belonging to them by using the sparsity-promoting prior J_s in (8). The algorithm details are presented next.

4.1. Outer Loop

The outer loop aims to detect the set of the active clusters, hence, we enforce $p(\mathbf{X})$ to promote the cluster-sparsity by $-\log p(\mathbf{X}) = \sum_{l=1}^{C} \log(\sum_{i \in \mathcal{C}_l} \|\mathbf{x}_i\| + \epsilon_0)$. Since $-\log p(\mathbf{X})$ is concave, we apply a majorization-minimization (MM) approximation to linearize $-\log p(\mathbf{X}) \approx \sum_{i=1}^{N} q_i^{(k_\text{c})} \|\mathbf{x}_i\|$, where $q_i^{(k_\text{c})} = \left(\sum_{i \in \mathcal{C}_l} \|\mathbf{x}_i^{(k_\text{c})}\| + \epsilon_0\right)^{-1}$ and k_c is the MM iteration index for the outer loop. Thus, the relaxed version of the problem (10) can be solved iteratively as

$$\{\hat{\mathbf{X}}^{(k_\text{c}+1)}, \boldsymbol{\Sigma}^{(k_\text{c}+1)}\} = \min_{\mathbf{X}, \boldsymbol{\Sigma}} \frac{1}{2}\|\mathbf{Y} - \boldsymbol{\Phi}\mathbf{X}^\text{T}\|_\text{F}^2 + \beta_1 \sum_{i=1}^{N} q_i^{(k_\text{c})} \|\mathbf{x}_i\|$$
$$+ \beta_2 \sum_{l=1}^{C} \sum_{i \in \mathcal{C}_l} \mathbf{x}_i^\text{H} \boldsymbol{\Sigma}_l \mathbf{x}_i - \sum_{l=1}^{C} \mu^{(k_\text{c})} \log|\boldsymbol{\Sigma}_l| + \beta_3 L \sum_{l=1}^{C} \text{tr}(\mathbf{B}_l^{-1}\boldsymbol{\Sigma}_l),$$
(11)

where $\mu^{(k_\text{c})} = (\beta_2 \sum_{i \in \mathcal{C}_l} p_i^M q_i^{(k_\text{c})} \|\mathbf{x}_i^{(k_\text{c})}\| + \beta_3 L d)$.

We develop a computationally efficient ADMM solution for (11) through a set of sequential update rules, each computed in closed-form. We introduce two splitting variables $\mathbf{Z}, \mathbf{V} \in \mathbb{C}^{M \times N}$ and the Lagrange dual variable matrices $\boldsymbol{\Lambda}_v, \boldsymbol{\Lambda}_z$ and define the set of variables to be estimated as $\boldsymbol{\Theta} = \{\mathbf{X}, \boldsymbol{\Sigma}, \mathbf{Z}, \mathbf{V}, \boldsymbol{\Lambda}_\text{z}, \boldsymbol{\Lambda}_\text{v}\}$. Subsequently, we write the augmented Lagrangian as

$$\mathcal{L}(\boldsymbol{\Theta}) = \frac{1}{2}\|\mathbf{Y} - \boldsymbol{\Phi}\mathbf{Z}^\text{T}\|_\text{F}^2 + \beta_1 \sum_{i=1}^{N} q_i^{(k_\text{c})} \|\mathbf{x}_i\| + \beta_2 \sum_{i=1}^{N} \mathbf{v}_i^\text{H} \boldsymbol{\Sigma}_i \mathbf{v}_i$$
$$- \sum_{l=1}^{C} \left(\mu^{(k_\text{c})} \log|\boldsymbol{\Sigma}_l| + \beta_3 L \text{tr}(\mathbf{B}_l^{-1}\boldsymbol{\Sigma}_l)\right) + \frac{\rho}{2}\|\mathbf{X} - \mathbf{V} + \frac{1}{\rho}\boldsymbol{\Lambda}_\text{v}\|_\text{F}^2$$
$$+ \frac{\rho}{2}\|\mathbf{X} - \mathbf{Z} + \frac{1}{\rho}\boldsymbol{\Lambda}_\text{z}\|_\text{F}^2 - \frac{\|\boldsymbol{\Lambda}_\text{z}\|_\text{F}^2}{2\rho} - \frac{\|\boldsymbol{\Lambda}_\text{v}\|_\text{F}^2}{2\rho}.$$
(12)

ADMM solves to the optimization problem (11) by minimizing the augmented Lagrangian $\mathcal{L}(\boldsymbol{\Theta})$ in (12) over the primal variables $(\mathbf{Z}, \mathbf{V}, \mathbf{X}, \boldsymbol{\Sigma})$, followed by updating the dual variables $(\boldsymbol{\Lambda}_\text{z}, \boldsymbol{\Lambda}_\text{v})$ [19]. Primal variable update is given by

$$\mathbf{Z}^{(k_\text{c}+1)} = \min_{\mathbf{Z}} \frac{1}{2}\|\boldsymbol{\Phi}\mathbf{Z}^\text{T} - \mathbf{Y}\|_\text{F}^2 + \frac{\rho}{2}\|\mathbf{X}^{(k_\text{c})} - \mathbf{Z} + \frac{1}{\rho}\boldsymbol{\Lambda}_\text{z}^{(k_\text{c})}\|_\text{F}^2 \quad (13)$$

$$\mathbf{V}^{(k_\text{c}+1)} = \min_{\mathbf{V}} \beta_2 \sum_{i=1}^{N} \mathbf{v}_i^\text{H} \boldsymbol{\Sigma}_l^{(k_\text{c})} \mathbf{v}_i + \frac{\rho}{2}\|\mathbf{X}^{(k_\text{c})} - \mathbf{V} + \frac{\boldsymbol{\Lambda}_\text{v}^{(k_\text{c})}}{\rho}\|_\text{F}^2 \quad (14)$$

$$\mathbf{X}^{(k_\text{c}+1)} = \min_{\mathbf{X}} \sum_{i=1}^{N} \alpha_i^{(k_\text{c})} \|\mathbf{x}_i\| + \frac{\rho}{2}\|\mathbf{X} - \mathbf{C}^{(k_\text{c}+1)}\|_\text{F}^2 \quad (15)$$

$$\boldsymbol{\Sigma}_l^{(k_\text{c}+1)} = \min_{\boldsymbol{\Sigma}_l} \beta_2 \sum_{i \in \mathcal{C}_l} \mathbf{v}_i^{(k_\text{c}+1)\text{H}} \boldsymbol{\Sigma}_l \mathbf{v}_i^{(k_\text{c}+1)} - \mu^{(k_\text{c}+1)} \log|\boldsymbol{\Sigma}_l|$$
$$+ \beta_3 L \text{tr}(\mathbf{B}_l^{-1}\boldsymbol{\Sigma}_l), \ l = 1, \ldots, C, \quad (16)$$

where $\mathbf{C}^{(k_\text{c})} = \frac{1}{2}(\mathbf{Z}^{(k_\text{c}+1)} + \mathbf{V}^{(k_\text{c}+1)} - \frac{\boldsymbol{\Lambda}_\text{z}^{(k_\text{c})} + \boldsymbol{\Lambda}_\text{v}^{(k_\text{c})}}{\rho})$ and $\alpha_i^{(k_\text{c})} = \left(\beta_1 q_i^{(k_\text{c})} - \beta_2 \log|\boldsymbol{\Sigma}_l^{(k_\text{c})}|q_i^{(k_\text{c})}\right)$.

4.2. Inner Loop

After running the outer loop for some pre-defined K_c iterations, we detect the set of the estimated active clusters $\hat{\mathcal{S}} = \bigcup_{j \in \mathcal{J}} \mathcal{C}_j$, where $l \in \mathcal{J}$ if there exists $i \in \mathcal{C}_l$ such that $\|\mathbf{x}_i\| > \epsilon$, where $\epsilon > 0$ is a small predefined parameter. In the inner loop, the proposed algorithm aims to detect the active UEs belonging to $\hat{\mathcal{S}}$ by using the separable sparsity-promoting prior $-\log p(\mathbf{X}) \propto \sum_{i \in \hat{\mathcal{S}}} \log(\|\mathbf{x}_i\| + \epsilon_0)$. Furthermore, we apply the MM approximation to linearize the concave $-\log p(\mathbf{X}) \approx \sum_{i \in \hat{\mathcal{S}}} g_i^{(k_\text{c})} \|\mathbf{x}_i\|$, where k_u is inner loop iteration index and $g_i^{(k_\text{u})} = \left(\|\mathbf{x}_i^{(k_\text{u})}\| + \epsilon_0\right)^{-1}$. The optimization problem for the inner loop is given by

$$\{\hat{\mathbf{X}}_{\hat{\mathcal{S}}}^{(k_\text{u}+1)}, \boldsymbol{\Sigma}^{(k_\text{u}+1)}\} = \min_{\mathbf{X}_{\hat{\mathcal{S}}}, \boldsymbol{\Sigma}} \frac{1}{2}\|\mathbf{Y} - \boldsymbol{\Phi}_{\hat{\mathcal{S}}} \mathbf{X}_{\hat{\mathcal{S}}}^\text{T}\|_\text{F}^2 + \beta_1 \sum_{i \in \hat{\mathcal{S}}} g_i^{(k_\text{u})} \|\mathbf{x}_i\|$$
$$+ \beta_2 \sum_{l \in \mathcal{J}} \sum_{i \in \hat{\mathcal{S}}_l} \mathbf{x}_i^\text{H} \boldsymbol{\Sigma}_l \mathbf{x}_i - \beta_2 \sum_{l \in \mathcal{J}} \log|\boldsymbol{\Sigma}_l|(\sum_{i \in \mathcal{C}_l} g_i^{(k_\text{u})} \|\mathbf{x}_i\| + \beta_3 L d)$$
$$+ \beta_3 L \sum_{l \in \mathcal{J}} \text{tr}(\mathbf{B}_l^{-1}\boldsymbol{\Sigma}_l),$$
(17)

where $\boldsymbol{\Phi}_{\hat{\mathcal{S}}}$ and $\mathbf{X}_{\hat{\mathcal{S}}}$ denote the matrices $\boldsymbol{\Phi}$ and $\mathbf{X}$, respectively, restricted to the set of the estimated active clusters, $\hat{\mathcal{S}}$.

Algorithm 1: The proposed JUICE algorithm

Input: $\boldsymbol{\Phi}$, $\{\mathbf{B}_l\}_{l=1}^C$, $\beta_1, \beta_2, \beta_3$, ρ, ϵ_0, ϵ, $k_{u_{max}}$, $k_{c_{max}}$, K_c.
Initialization: $\mathbf{X}^{(0)}$, $\mathbf{V}^{(0)}$, $\mathbf{Z}^{(0)}$, $\boldsymbol{\Lambda}_v^{(0)}$, $\boldsymbol{\Lambda}_z^{(0)}$, $k_u = 1$, $k_c = 1$.

1 BS receives $\mathbf{Y}$, compute and store $(\boldsymbol{\Phi}^T \boldsymbol{\Phi}^* + \rho \mathbf{I}_N)^{-1}$
2 **while** $k_c < k_{c_{max}}$ or $\|\mathbf{X}^{(k_c)} - \mathbf{X}^{(k-1)}\| < \epsilon$ **do**
3    $\mathbf{Z}^{(k_c+1)} = (\rho \mathbf{X}^{(k_c)} + \boldsymbol{\Lambda}^{(k_c)} + \mathbf{Y}^T \boldsymbol{\Phi}^*)(\boldsymbol{\Phi}^T \boldsymbol{\Phi}^* + \rho \mathbf{I}_N)$
4    $\mathbf{v}_i^{(k_c+1)} = (\beta_2 \boldsymbol{\Sigma}_i^{(k_c)} + \rho \mathbf{I}_M)^{-1}(\rho \mathbf{x}_i^{(k)} + \boldsymbol{\lambda}_{vi}^{(k_c)})$, $i = 1, \ldots, N$
5    $\mathbf{x}_i^{(k_c+1)} = \dfrac{\max\{0, \|\mathbf{c}_i^{(k_c)}\| - \frac{\alpha_i^{(k_c)}}{2\rho}\} \mathbf{c}_i^{(k_c)}}{\|\mathbf{c}_i^{(k_c)}\|}$, $i = 1, \ldots, N$
6    $\boldsymbol{\Sigma}_l^{(k_c)+1} = (\beta_2 \sum_{i \in \mathcal{C}_l} \mathbf{v}_i \mathbf{v}_i^H + \beta_3 L \mathbf{B}_l^{-1}) \mu^{(k_c)}$, $l = 1, \ldots, C$
7    $\boldsymbol{\Lambda}_z^{(k_c+1)} = \boldsymbol{\Lambda}_z^{(k_c)} + \rho(\mathbf{X}^{(k_c+1)} - \mathbf{Z}^{(k_c+1)})$
8    $\boldsymbol{\Lambda}_v^{(k_c+1)} = \boldsymbol{\Lambda}_v^{(k_c)} + \rho(\mathbf{X}^{(k_c+1)} - \mathbf{V}^{(k_c+1)})$
9   **if** $(k_c \bmod K_c) = 0$ **then**
10      $\hat{\mathcal{S}} = \cup_{j \in \mathcal{J}} \mathcal{C}_j$, $\{l \in \mathcal{J} : \exists i \in \mathcal{C}_l | \|\mathbf{x}_i\| > \epsilon\}$
11   **while** $k_u < k_{u_{max}}$ **do**
12   Solve (17) using the similar update rules as (18), but
         using $g_i^{(k_u)} = (\|\mathbf{x}_i^{(k_u)} + \epsilon_0\|)^{-1}$
13        $k_u \leftarrow k_u + 1$
14      $\mathbf{X}_{\hat{\mathcal{S}}}^{(k_c)} = \mathbf{X}_{\hat{\mathcal{S}}}^{(k_u)}$, $\mathbf{Z}_{\hat{\mathcal{S}}}^{(k_c)} = \mathbf{Z}_{\hat{\mathcal{S}}}^{(k_u)}$, $\mathbf{V}_{\hat{\mathcal{S}}}^{(k_c)} = \mathbf{V}_{\hat{\mathcal{S}}}^{(k_u)}$,
     $\boldsymbol{\Lambda}_{z\hat{\mathcal{S}}}^{(k_c)} = \boldsymbol{\Lambda}_{z\hat{\mathcal{S}}}^{(k_u)}$, $\boldsymbol{\Lambda}_{v\hat{\mathcal{S}}}^{(k_c)} = \boldsymbol{\Lambda}_{v\hat{\mathcal{S}}}^{(k_u)}$
15      $k_u = 1$,
16    $k_c \leftarrow k_c + 1$

Fig. 1: Performance evaluation with clustered activity pattern.

Fig. 2: Performance evaluation with random activity pattern.

4.3. ADMM Update Rules

Owing to the applied splitting techniques, all the sub-problems (13)–(16) are convex, thus, can be solved analytically via closed-form formulas. Further, the optimization over $\mathbf{X}$, $\mathbf{V}$, and $\boldsymbol{\Sigma}$ is separable over the UEs and the clusters, allowing for parallel updates. Thus, the outer loop update is given as

$$\begin{cases} \mathbf{Z}^{(k_c+1)} = (\rho \mathbf{X}^{(k_c)} + \boldsymbol{\Lambda}^{(k_c)} + \mathbf{Y}^T \boldsymbol{\Phi}^*)(\boldsymbol{\Phi}^T \boldsymbol{\Phi}^* + \rho \mathbf{I}_N)^{-1}, \\ \mathbf{v}_i^{(k_c+1)} = (\beta_2 \boldsymbol{\Sigma}_i^{(k_c)} + \rho \mathbf{I}_M)^{-1}(\rho \mathbf{x}_i^{(k)} + \boldsymbol{\lambda}_{vi}^{(k_c)}), \quad \forall i \in \mathcal{N}, \\ \mathbf{x}_i^{(k_c+1)} = \max\{0, \|\mathbf{c}_i^{(k_c)}\| - \frac{\alpha_i^{(k_c)}}{2\rho}\} \|\mathbf{c}_i^{(k_c)}\| \mathbf{c}_i^{(k)}, \quad \forall i \in \mathcal{N}, \\ \boldsymbol{\Sigma}_l^{(k_c)+1} = (\beta_2 \sum_{i \in \mathcal{C}_l} \mathbf{v}_i^{(k_c)} \mathbf{v}_i^{(k_c)H} + \beta_3 L \mathbf{B}_l^{-1}) \mu^{(k_c)}, l = 1, \ldots, C. \end{cases}$$
(18)

The update rules for the inner loop, i.e., the solution to (17) is similar to (18) but with changing $q_i^{(k_c)}$ to $g_i^{(k_u)}$. Algorithm 1 provides the proposed ADMM solution to optimization problem (10).

5. NUMERICAL RESULTS

Consider a network with one BS equipped with $M = 20$ antennas serving a $N = 500$ UEs distributed equally over $C = 20$ clusters, where $K = 16$ UEs are active at each T_c. Each UE is assigned a unique unit-norm pilot sequence that is generated from an i.i.d. complex Bernoulli distribution. We set each $\mathbf{B}_l, \forall l$, as $\mathbf{B}_l = \zeta \boldsymbol{\Psi}_l + (1 - \zeta) \frac{1}{L} \sum_{i \in \mathcal{C}_l} \boldsymbol{\Sigma}_i$, where $\boldsymbol{\Psi}_l$ is a random semi-definite Hermitian matrix to model the error in the prior knowledge on the true precision matrix $\boldsymbol{\Sigma}_l$, whereas the parameter ζ controls the level of average mismatch between $\boldsymbol{\Sigma}_i$ and $\mathbf{B}_l, \forall i \in \mathcal{C}_l$, and it is set as $\zeta = 0.1$.

Channel estimation is quantified in terms of normalized mean square error (NMSE) defined as $\frac{\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}_{\mathcal{S}}\|_F^2]}{\mathbb{E}[\|\mathbf{X}\|_F^2]}$, where the expectation is computed via Monte-Carlo averaging over all sources of randomness. UEs activity detection is quantified in terms of support recovery rate (SRR), defined as $\frac{|\mathcal{S} \cap \hat{\mathcal{S}}|}{|\mathcal{S} - \hat{\mathcal{S}}| + K}$, where $\hat{\mathcal{S}}$ denotes the de-
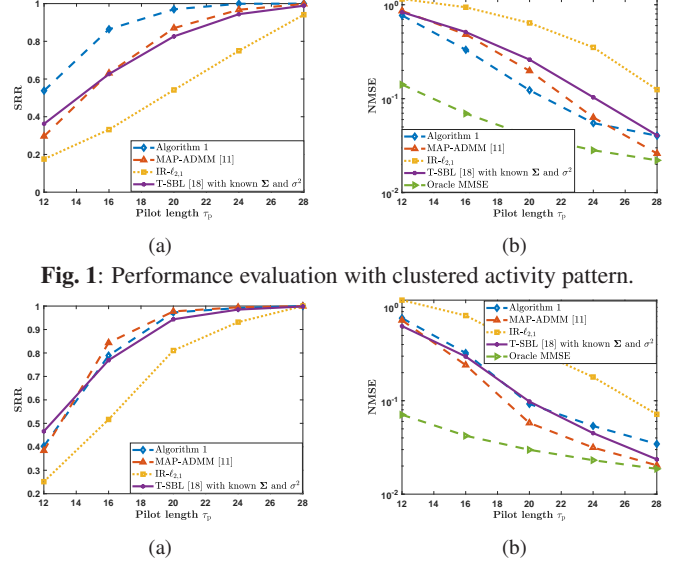
tected support by a given algorithm.

The performance of proposed Algorithm 1 is compared with: 1) iterative reweighted $\ell_{2,1}$ (IR-$\ell_{2,1}$) via ADMM [10], 2) MAP-ADMM [11], and 3) T-SBL [18], where for MAP-ADMM and T-SBL, both $\boldsymbol{\Sigma}$ and σ^2 are known to the BS. For an optimal NMSE benchmark, we consider the oracle minimum mean square error (MMSE) estimator having prior knowledge on the true set of active UEs and their covariance matrices.

We consider two cases for UEs' activity pattern: 1) clustered: with 2 active clusters, each with 8 active UEs, and 2) random, where $K = 16$ active UEs are randomly distributed among the clusters.

Fig. 1(a) presents the SRR against the pilot length for the clustered activity pattern. The result shows that Algorithm 1 provides the highest SSR, which can be explained by the fact that Algorithm 1 effectively exploits the clustered structure of the activity pattern using the proposed prior functions. Similarly, the proposed algorithm moderately outperforms MAP-ADMM and T-SBL in terms of NMSE (Fig. 1(b)). In summary, Fig. 1 clearly shows the utility of the proposed algorithm in JUICE with clustered activity patterns.

Fig. 2(a) illustrates the SRR for the random activation scenario. Algorithm 1 achieves performance similar to MAP-ADMM and T-SBL while providing significant gains over IR-$\ell_{2,1}$. On the other hand, Fig. 2(b) shows that while the proposed algorithm clearly outperforms IR-$\ell_{2,1}$ in terms of channel estimation accuracy, it provides just a slightly inferior NMSE performance compared to T-SBL and MAP-ADMM. These results show the robustness of the proposed algorithm to the structure of the activity pattern: the pattern can deviate from the desired clustered form to fully random activity, rendering the algorithm effective for a wide range of UE activity models.

6. CONCLUSION

We derived a Bayesian framework to address the JUICE with clustered activity in mMTC under correlated MIMO channels without prior knowledge of the exact CDI. To encourage solutions with cluster sparsity, we proposed a cluster-sparsity-promoting function that correlates the activity of the UEs belonging to the same cluster. We developed an ADMM algorithm that provides a computationally-efficient solution via a sequence of closed-form update rules. The numerical results showed the improvement and robustness brought by the proposed algorithm.

7. REFERENCES

[1] Zhilin Chen, Foad Sohrabi, and Wei Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Processing*, vol. 66, no. 7, pp. 1890–1904, 2018.

[2] Liang Liu and Wei Yu, "Massive connectivity with massive MIMO—part I: Device activity detection and channel estimation," *IEEE Trans. Signal Processing*, vol. 66, no. 11, pp. 2933–2946, 2018.

[3] Liang Liu and Wei Yu, "Massive connectivity with massive MIMO—part II: Achievable rate characterization," *IEEE Trans. Signal Processing*, vol. 66, no. 11, pp. 2947–2959, 2018.

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

[5] Malong Ke, Zhen Gao, Yongpeng Wu, Xiqi Gao, and Robert Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Trans. Signal Processing*, vol. 68, pp. 764–779, 2020.

[6] Xiaoxu Zhang, Ying-Chang Liang, and Jun Fang, "Novel Bayesian inference algorithms for multiuser detection in M2M communications," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7833–7848, 2017.

[7] Luca Sanguinetti, Emil Björnson, and Jakob Hoydis, "Towards massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, 2020.

[8] Yiyao Cheng, Lei Liu, and Li Ping, "Orthogonal AMP for massive access in channels with spatial and temporal correlations," *IEEE J. Select. Areas Commun.*, vol. 39, no. 3, pp. 726–740, 2021.

[9] Jianan Bai and Erik G. Larsson, "Activity detection in distributed mimo: Distributed amp via likelihood ratio fusion," *IEEE Wireless Commun. Lett*, pp. 1–1, 2022.

[10] Hamza Djelouat, Markus Leinonen, and Markku Juntti, "Iterative reweighted algorithms for joint user identification and channel estimation in spatially correlated massive mtc," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 4805–4809.

[11] Hamza Djelouat, Markus Leinonen, and Markku Juntti, "Spatial correlation aware compressed sensing for user activity detection and channel estimation in massive MTC," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6402–6416, 2022.

[12] H. Djelouat, M. Leinonen, L. Ribeiro, and M. Juntti, "Joint user identification and channel estimation via exploiting spatial channel covariance in mMTC," *IEEE Wireless Commun. Lett*, vol. 10, no. 4, pp. 887–891, 2021.

[13] Emil Björnson, Jakob Hoydis, and Luca Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[14] Emil Björnson, Luca Sanguinetti, and Merouane Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. Annual Asilomar Conf. Signals, Syst., Comp.*, 2016, pp. 974–978.

[15] Jongmin Kim, Woohyuk Chang, Bangchul Jung, Dror Baron, and Jong Chul Ye, "Belief propagation for joint sparse recovery," *arXiv preprint arXiv:1102.3289*, 2011.

[16] Christian Steffens, Marius Pesavento, and Marc E Pfetsch, "A compact formulation for the $l_{2,1}$ mixed-norm minimization problem," *IEEE Trans. Signal Processing*, vol. 66, no. 6, pp. 1483–1497, 2018.

[17] Christopher M Bishop and Nasser M Nasrabadi, *Pattern recognition and machine learning*, vol. 4, 2006.

[18] Zhilin Zhang and Bhaskar D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Select. Topics Signal Processing*, vol. 5, no. 5, pp. 912–926, 2011.

[19] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.