Joint User Identification and Channel Estimation via Exploiting Spatial Channel Covariance in mMTC

Hamza Djelouat, Markus Leinonen, *Member, IEEE*, Lucas Ribeiro, *Graduate Student Member, IEEE*, and Markku Juntti, *Fellow, IEEE*

***Abstract*—Grant-free random access is a key enabler in massive machine-type communications (mMTC) to reduce signalling overhead and latency thereby improving the energy efficiency. One of its main challenges lies in joint user activity identification and channel estimation (JUICE). Due to the sporadic mMTC traffic, JUICE can be solved as a compressive sensing (CS) problem. We address CS-based JUICE in uplink with single-antenna transmitters and a multiantenna base station under spatially correlated fading channels. We formulate a novel CS problem that utilizes prior information on the second order statistics of the channel of each user to improve the performance. We propose a method based on alternating direction method of multipliers to solve the JUICE efficiently. The simulation results show that the proposed method significantly improves the user identification accuracy and channel estimation performance with lower signalling overhead as compared to the baseline schemes.**

***Index Terms*—ADMM, channel estimation, mMTC, spatially correlated channels, user detection.**

I. Introduction

MASSIVE machine-type communications (mMTC) aim to provide wireless connectivity to billions of low-cost energy-constrained Internet of Things (IoT) devices [1]. IoT and mMTC have three main features different to more classical personal communications of mobile phones, for example. First, transmission is sporadic, i.e., only a small and varying subset of IoT devices is active at a given transmission instant. Second, they have short-packet transmission dominated by uplink traffic. Third, energy-efficient communication protocols are used to ensure a long lifespan for the IoT devices. On the other hand, one of the main challenges for mMTC is the channel access. Since the signalling overheads are proportional to the number of devices, grant-based random access protocols are not able to support massive connectivity.

Grant-free multiple-access has been identified as a key enabler for mMTC [2]. In grant-free access, the IoT devices transmit data as per their needs without going through channel access protocols. The main advantage of grant-free access compared to conventional random access is the reduced signalling overhead and the improved energy-efficiency of the IoT devices. However, the main challenge in grant-free access is to detect the set of active users and to estimate their channels for coherent detection, conventionally known as joint user identification and channel estimation (JUICE).

In mMTC, the sparsity due to the sporadic user activation motivates to formulate the JUICE as a compressed sensing (CS) [3] problem. In particular, as the BS antennas sense the same sparse user activity, the JUICE problem extends to the multiple measurement vector (MMV) CS framework. Chen *et al.* [4] and Liu and Yu [5] utilized approximate message passing for JUICE and provided asymptotic analysis on user activity detection and channel estimation performance. Senel and Larsson [6] designed a "non-coherent" detection scheme for very short packet transmission by jointly detecting the active users and the transmitted information bits. A maximum likelihood estimation approach using the measurement covariance matrix was considered in [7]. The vast majority of JUICE related works assume that the communication channels are spatially uncorrelated. Although this assumption leads to analytically tractable solutions, it is not always the case in practice [8].

We address the JUICE problem in mMTC under spatially correlated fading channels. There are two main reasons to consider such channel model: 1) the practical massive multiple-input multiple-output (MIMO) channels are essentially always spatially correlated [8], and 2) spatial correlation can be very beneficial for the multi-user MIMO systems if the devices have sufficiently different spatial correlation matrices [8], [9]. In addition, as the channel spatial correlation varies in a slower time-scale compared to the channel realizations, the channel covariance matrices for all IoT devices can be estimated with high accuracy in practice [10]. Therefore, the channel information inherently embedded within the spatial correlation matrices can be used to enhance the JUICE performance.

The main contributions of this letter are twofold. First, we consider spatially correlated fading channels and formulate a novel JUICE problem to exploit the prior information on the second order channel statistics. Second, we propose a computationally efficient method based on alternating direction method of multipliers (ADMM) to solve the problem. The proposed approach is empirically shown to significantly enhance the user activity detection accuracy and, consequently, the channel estimation performance. Moreover, the proposed approach achieves the same performance as baseline MMV JUICE, yet with a smaller signalling overhead.

II. System Model and Problem Formulation

A. mMTC Uplink Channel Model

Consider a single cell uplink communication scheme, shown in Fig. 1, consisting of a set of N' uniformly distributed single-antenna users $\mathcal{N} = \{1, \ldots, N\}$ and a single BS

Manuscript received November 10, 2020; accepted December 28, 2020. Date of publication January 5, 2021; date of current version April 9, 2021. This work was supported in part by the Academy of Finland under Grant 319485, and in part by the Academy of Finland 6Genesis Flagship under Grant 318927. The work of Markus Leinonen was supported in part by the Infotech Oulu, and in part by the Academy of Finland under Grant 323698. The associate editor coordinating the review of this article and approving it for publication was W. Cheng. *(Corresponding author: Hamza Djelouat.)*

The authors are with the Center for Wireless Communications—Radio Technologies, University of Oulu, Oulu 90014, Finland (e-mail: hamza.djelouat@oulu.fi; markus.leinonen@oulu.fi; lucas.ribeiro@oulu.fi; markku.juntti@oulu.fi).

Digital Object Identifier 10.1109/LWC.2021.3049167

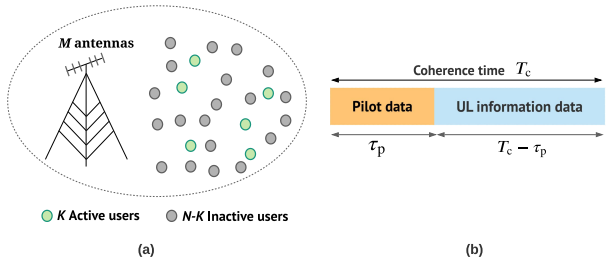

Fig. 1. (a) mMTC uplink scenario with sporadic user activation, (b) Division of a coherence interval T_c.

equipped with a uniform linear array (ULA) containing M antennas.

We consider a block fading channel response over each coherence period T_c. We further assume that the BS is located in an elevated position and most of the channel power lies in a limited number of spatial directions [10]. The users usually have a rich scattering environment around them. Therefore, the received signal from each user $i \in \mathcal{N}$ is the superposition of P_i physical signal paths, each reaching the BS as a plane wave. Accordingly, the channel response vector $\mathbf{h}_i \in \mathbb{C}^M$ between the ith user and the BS is modelled as

$$\mathbf{h}_i = \frac{1}{\sqrt{P_i}} \sum_{p=1}^{P_i} g_{i,p} \mathbf{a}(\psi_{i,p}), \tag{1}$$

where $g_{i,p} \in \mathbb{C}$ accounts for the gain of the pth propagation path for the ith user. The steering vector $\mathbf{a}(\psi_{i,p}) \in \mathbb{C}^M$ is the array response of the ULA, which is given by $[\mathbf{a}(\psi_{i,p})]_m = e^{-j2\pi(m-1)\Delta_r \cos(\psi_{i,p})}$, where Δ_r denotes the normalized spacing between the adjacent antenna elements at the receiver [11], and $\psi_{i,p}$ is the angle of arrival (AoA) of the pth path for user i. Each $\psi_{i,p}$ follows a uniform distribution $\psi_{i,p} \sim U(\bar{\psi}_i - \frac{\zeta}{2}, \bar{\psi}_i + \frac{\zeta}{2})$, where $\bar{\psi}_i \in [-\pi/2, \pi/2]$ represents the incident angle between the ith user and the BS, and ζ denotes the angular spread of scatterers around the users.

The channel realizations $\mathbf{h}_i$ in (1) are assumed to be independent between different coherence intervals T_c. We consider users with low mobility and adopt the common assumption that the channel is wide-sense stationary [10], i.e., the channel covariance matrix of user i, denoted as $\mathbf{R}_i = \mathbb{E}[\mathbf{h}_i \mathbf{h}_i^{\mathrm{H}}] \in \mathbb{C}^{M \times M}$, varies in a slower time-scale compared to the channel realizations [12]. Accordingly, $\{\mathbf{R}_i\}_{i=1}^N$ are assumed to remain fixed for τ_s coherence intervals, where τ_s can be on the order of thousands [8] and $\{\mathbf{R}_i\}_{i=1}^N$ are known to the BS [10]. Nevertheless, in practice, the BS does not know the exact channel covariance matrices $\{\mathbf{R}_i\}_{i=1}^N$, rather, during a training phase, the BS acquires an estimate[1] of the second-order statistics of the channel, $\{\hat{\mathbf{R}}_i\}_{i=1}^N$. Thus, for each user $i \in \mathcal{N}$, the BS obtains T estimates of channel response $\mathbf{h}_i$ from different coherence intervals, denoted as $\hat{\mathbf{h}}_i^1, \ldots, \hat{\mathbf{h}}_i^T$. We define the estimated channel covariance for each user i as $\hat{\mathbf{R}}_i = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{h}}_i^t \hat{\mathbf{h}}_i^{t\,\mathrm{H}}$.

Due to the sporadic nature of mMTC, at each coherence interval T_c, only $K < N$ users are active. For coherent data detection, the active users have to be detected and their channels have to be estimated. To this end, the BS assigns to each user $i \in \mathcal{N}$ a unit-norm pilot sequence $\boldsymbol{\phi}_i \in \mathbb{C}^{\tau_p}$. During each T_c, K active users transmit their pilot sequences to the BS. To mitigate the channel gain difference between the users, a power control policy is deployed such that user i transmits with a power ρ_i^{UL} that is inversely proportional to the average channel gain [6], [8]. Accordingly, the received pilot signal $\mathbf{Y} \in \mathbb{C}^{\tau_p \times M}$ at the BS is given by

$$\mathbf{Y} = \sum_{i=1}^{N} \gamma_i \sqrt{\rho_i^{\mathrm{UL}}} \boldsymbol{\phi}_i \mathbf{h}_i^{\mathrm{T}} + \mathbf{W}, \tag{2}$$

where $\mathbf{W} \in \mathbb{C}^{\tau_p \times M}$ is additive white Gaussian noise with independent and identically distributed (i.i.d) elements as $\mathcal{CN}(0, \sigma^2)$, and $\gamma_i \in \mathbb{B}$ is a binary user activity indicator as

$$\gamma_i = \begin{cases} 1, & i \in \mathcal{S} \\ 0, & \text{otherwise}, \end{cases}$$

where $\mathcal{S} \subseteq \{1, \ldots, N\}$, $|\mathcal{S}| = K$, is the set of active users.

We define the effective channel of user $i \in \mathcal{N}$ as $\mathbf{x}_i = \gamma_i \sqrt{\rho_i^{\mathrm{UL}}} \mathbf{h}_i$ and the effective channel matrix as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{C}^{M \times N}$. Let $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_N] \in \mathbb{C}^{\tau_p \times N}$ be the pilot matrix. We can rewrite (2) as

$$\mathbf{Y} = \boldsymbol{\Phi} \mathbf{X}^{\mathrm{T}} + \mathbf{W}. \tag{3}$$

B. Problem Definition

1) Standard JUICE: The columns of the effective channel matrix $\mathbf{X}$ corresponding to the inactive users are zero, hence, $\mathbf{X}^{\mathrm{T}}$ has a *row-sparse* structure. Therefore, joint user identification and channel estimation reduces to the detection of the support of $\mathbf{X}^{\mathrm{T}}$ (i.e., the non-zero columns of $\mathbf{X}$) as well as estimating their coefficients. Thus, JUICE can be modeled as a sparse MMV reconstruction problem. Algorithms developed to solve MMV problems include mixed norm minimization [13] (and the references therein), alternating direction methods [14], and sparse Bayesian learning [15].

We formulate the JUICE as an $\ell_{2,1}$-norm minimization problem

$$\min_{\mathbf{X}} \frac{1}{2} \|\boldsymbol{\Phi} \mathbf{X}^{\mathrm{T}} - \mathbf{Y}\|_{\mathrm{F}}^2 + \beta_1 \|\mathbf{X}^{\mathrm{T}}\|_{2,1}, \tag{4}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, $\ell_{2,1}$-norm is defined as $\|\mathbf{X}^{\mathrm{T}}\|_{2,1} = \sum_{i=1}^N \|\mathbf{x}_i\|_2$, and parameter β_1 introduces a trade-off between the measurement fidelity $\|\boldsymbol{\Phi} \mathbf{X}^{\mathrm{T}} - \mathbf{Y}\|_{\mathrm{F}}^2$ and the sparsity level in $\mathbf{X}$.

2) JUICE With Channel Covariance Information: The problem in (4) utilizes only the sparse structure of effective channel matrix $\mathbf{X}$, and it does not exploit the information embedded in the estimated channel covariance matrices $\{\hat{\mathbf{R}}_i\}_{i=1}^N$. On this account, we propose a novel JUICE problem formulation that exploits both the sparsity and the covariance information available at the BS. The key idea is that since the channel covariance matrices vary very slowly, the sample covariance matrix $\mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}$ for the effective channel computed over a single coherence time for each active user $i \in \mathcal{S}$ should carry similar information as the estimated covariance matrix $\hat{\mathbf{R}}_i$ that have been computed by the BS in the training phase.

Based on the above argument, we augment the optimization problem in (4) with a regularization term that penalizes the deviation of the sample covariance matrix $\mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}$ from the

[1]Channel covariance estimation in massive MIMO networks is a separate topic and it is out of the scope of this letter; for more references, see [12].

estimated scaled covariance matrix $\tilde{\mathbf{R}}_i = \rho_i^{\mathrm{UL}}\hat{\mathbf{R}}_i$. Thus, the modified JUICE problem that exploits the prior information on the second order channel statistics is expressed as

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{\Phi}\mathbf{X}^{\mathrm{T}} - \mathbf{Y}\|_{\mathrm{F}}^2 + \beta_1\|\mathbf{X}\|_{2,1} + \beta_2\sum_{i=1}^{N}\mathbb{I}(\mathbf{x}_i)\|\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}} - \tilde{\mathbf{R}}_i\|_{\mathrm{F}}^2, \tag{5}$$

where the parameters β_1 and β_2 control the trade-off between the emphasis on the measurement consistency term, the sparsity-promoting term, and the covariance deviation term; the indicator function $\mathbb{I}(\cdot)$ is defined as

$$\mathbb{I}(\mathbf{x}_i) = \begin{cases} 1, & \|\mathbf{x}_i\|_2 > 0, \\ 0, & \|\mathbf{x}_i\|_2 = 0, \end{cases} \quad \forall i \in \mathcal{N}, \tag{6}$$

which ensures that only the estimated active users are penalized on the deviation with respect to the channel covariance.

The indicator function in (6) implies a *combinatorial* nature of the problem, and is, thus, hard to handle. We relax the indicator function with a smooth approximation $\mathbb{I}(\mathbf{x}_i) \approx \|\mathbf{x}_i\|_2$. This is a viable option: through the power control, the effective channels $\mathbf{x}_i$, $\forall i \in \mathcal{N}$, have the same range so the (magnitude-dependent) ℓ_2-norm does not cause bias among the estimated active users. Moreover, it allows a closed-form solution to a sub-problem involved in the optimization procedure, as shown later. By the relaxation of $\mathbb{I}(\mathbf{x}_i)$, we rewrite (5) as

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{\Phi}\mathbf{X}^{\mathrm{T}} - \mathbf{Y}\|_{\mathrm{F}}^2 + \beta_1\|\mathbf{X}\|_{2,1} + \beta_2\sum_{i=1}^{N}\|\mathbf{x}_i\|_2\|\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}} - \tilde{\mathbf{R}}_i\|_{\mathrm{F}}^2. \tag{7}$$

III. Efficient Solution via ADMM

In this section, we propose an efficient algorithm based on the alternating direction method of multipliers [16] to solve the optimization problem in (7). ADMM is a rather simple yet efficient framework that is well suited for high-dimensional optimization problems. For instance, Deng *et al.* [17] reported that ADMM outperforms coordinate gradient descent for sparse recovery from single measurement vector in terms of both signal recovery quality and convergence rate.

By introducing a set of auxiliary variables $\mathbf{Z} = \{\mathbf{Z}_i \in \mathbb{C}^{M\times M}\}_{i=1}^{N}$, we write the scaled form of augmented Lagrangian of (7) as

$$\mathcal{L}(\mathbf{X},\mathbf{Z},\mathbf{\Lambda}) = \beta_1\|\mathbf{X}\|_{2,1} + \frac{1}{2}\|\mathbf{\Phi}\mathbf{X}^{\mathrm{T}} - \mathbf{Y}\|_{\mathrm{F}}^2 - \frac{\rho}{2}\|\frac{\mathbf{\Lambda}_i}{\rho}\|_{\mathrm{F}}^2 + \beta_2\sum_{i=1}^{N}\|\mathbf{x}_i\|_2\|\mathbf{Z}_i - \tilde{\mathbf{R}}_i\|_{\mathrm{F}}^2 + \frac{\rho}{2}\|\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}} - \mathbf{Z}_i + \frac{1}{\rho}\mathbf{\Lambda}_i\|_{\mathrm{F}}^2. \tag{8}$$

where $\mathbf{\Lambda} = \{\mathbf{\Lambda}_i \in \mathbb{C}^{M\times M}\}_{i=1}^{N}$ denotes the set of Lagrangian multipliers and ρ is a positive ADMM penalty parameter.

The basic idea of ADMM is to iterate the optimization through three steps sequentially. First, $\mathcal{L}(\mathbf{X},\mathbf{Z},\mathbf{\Lambda})$ is minimized over the set of primal variables $\mathbf{Z}$ for fixed $\mathbf{X}$ and $\mathbf{\Lambda}$. Second, $\mathcal{L}(\mathbf{X},\mathbf{Z},\mathbf{\Lambda})$ is minimized over primal variable $\mathbf{X}$ for fixed $\mathbf{Z}$ and $\mathbf{\Lambda}$. Finally, we update the set of Lagrange multipliers $\mathbf{\Lambda}$ using a gradient ascent step. Thus, the ADMM iterates as follows [16, eqs. (3.2)–(3.4)]:

$$\mathbf{Z}_i^{(k+1)} := \min_{\mathbf{Z}_i}\Big\{\beta_2\|\mathbf{x}_i^{(k)}\|_2\|\mathbf{Z}_i - \tilde{\mathbf{R}}_i\|_{\mathrm{F}}^2 + \frac{\rho}{2}\|\mathbf{x}_i^{(k)}\mathbf{x}_i^{(k)\,\mathrm{H}} - \mathbf{Z}_i + \frac{1}{\rho}\mathbf{\Lambda}_i^{(k)}\|_{\mathrm{F}}^2\Big\},\ \forall i \in \mathcal{N}, \tag{9}$$

$$\mathbf{X}^{(k+1)} := \min_{\mathbf{X}}\Big\{\sum_{i=1}^{N}\alpha_i^{(k)}\|\mathbf{x}_i\|_2 + \frac{1}{2}\|\mathbf{\Phi}\mathbf{X}^{\mathrm{T}} - \mathbf{Y}\|_{\mathrm{F}}^2 + \frac{\rho}{2}\sum_{i=1}^{N}\|\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}} - \mathbf{Z}_i^{(k+1)} + \frac{1}{\rho}\mathbf{\Lambda}_i^{(k)}\|_{\mathrm{F}}^2\Big\}, \tag{10}$$

where $\alpha_i^{(k)} = (\beta_1 + \beta_2\|\mathbf{Z}_i^{(k+1)} - \tilde{\mathbf{R}}_i\|_{\mathrm{F}}^2)$, $\forall i \in \mathcal{N}$, and

$$\mathbf{\Lambda}_i^{(k+1)} := \mathbf{\Lambda}_i^{(k)} + \rho\Big(\mathbf{x}_i^{(k+1)}\mathbf{x}_i^{(k+1)\,\mathrm{H}} - \mathbf{Z}_i^{(k+1)}\Big),\ \forall i \in \mathcal{N}, \tag{11}$$

where (k) denotes the iteration index. We present the derivations of the ADMM steps (9) and (10) in detail below.

a) ***Z**-Update:* Note that finding the optimal solution for (9) decouples into solving N convex sub-problems in parallel. Thus, the unique optimal solution is obtained by setting the derivative to zero, resulting in

$$\mathbf{Z}_i^{(k+1)} = \frac{2\beta_2\|\mathbf{x}_i^{(k)}\|_2\tilde{\mathbf{R}}_i + \mathbf{\Lambda}_i^{(k)} + \rho\mathbf{x}_i^{(k)}\mathbf{x}_i^{(k)\,\mathrm{H}}}{2\beta_2\|\mathbf{x}_i^{(k)}\|_2 + \rho},\ \forall i \in \mathcal{N}. \tag{12}$$

b) ***X**-Update:* First, note that the third term in (10) is a fourth-order polynomial, thus, the objective function is non-convex. One way to solve (10) is to utilize standard iterative approaches such as the steepest descent methods. However, this will result in an additional inner iteration loop to be performed inside each ADMM iteration (k), rendering the solution to be inefficient for large-scale problems. Therefore, instead of solving (10) exactly, we utilize the Taylor expansion and *approximate* the solution to (10). The aim of using Taylor expansion is *threefold*. First, to linearize the quadratic term and the non-convex term. Second, to decompose the optimization problem into N separate sub-problems. Moreover, each of these N sub-problems admits a closed-form solution.

The quadratic term in (10) is approximated by the first-order Taylor expansion as

$$\|\mathbf{\Phi}\mathbf{X}^{\mathrm{T}} - \mathbf{Y}\|_{\mathrm{F}}^2 \approx \|\mathbf{\Phi}\mathbf{X}^{(k)} - \mathbf{Y}\|_{\mathrm{F}}^2 + \langle\mathbf{G}^{(k)}, \mathbf{X} - \mathbf{X}^{(k)}\rangle, \tag{13}$$

where $\mathbf{G}^{(k)} = 2(\mathbf{X}^{(k)}\mathbf{\Phi}^* - \mathbf{Y})\mathbf{\Phi}^{\mathrm{T}}$ is the gradient of $\|\mathbf{\Phi}\mathbf{X}^{\mathrm{T}} - \mathbf{Y}\|_{\mathrm{F}}^2$ with respect to $\mathbf{X}$ at point $\mathbf{X}^{(k)}$.

Similarly, the non-convex term in (10) is approximated by the first-order Taylor expansion as

$$\|\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}} - \mathbf{Z}_i^{(k+1)} + \frac{1}{\rho}\mathbf{\Lambda}_i\|_{\mathrm{F}}^2 \approx \|\mathbf{x}_i^{(k)}\mathbf{x}_i^{(k)\,\mathrm{H}} - \mathbf{Z}_i^{(k+1)} + \frac{1}{\rho}\mathbf{\Lambda}_i\|_{\mathrm{F}}^2 + \langle\mathbf{f}_i^{(k)}, \mathbf{x}_i - \mathbf{x}_i^{(k)}\rangle,\ \forall i \in \mathcal{N}, \tag{14}$$

where $\mathbf{f}_i^{(k)} = 4(\mathbf{x}_i^{(k)}\mathbf{x}_i^{(k)\,\mathrm{H}} - \mathbf{Z}_i^{(k+1)} + \frac{1}{\rho}\mathbf{\Lambda}_i^{(k)})\mathbf{x}_i$.

Due to the linearization, the Taylor approximation in (13) and (14) are accurate only for $\mathbf{X}$ near $\mathbf{X}^{(k)}$. Thus, we add a quadratic proximal regularization term $\frac{1}{\tau}\|\mathbf{X} - \mathbf{X}^{(k)}\|_{\mathrm{F}}^2$ with parameter $\tau > 0$ in the objective function of (10) [18]. Subsequently, we reformulate (10) as follows

$$\mathbf{X}^{(k+1)} \approx \min_{\mathbf{X}}\Big\{\sum_{i}^{N}\alpha_i^{(k)}\|\mathbf{x}_i\|_2 + \frac{1}{2}\langle\mathbf{G}^{(k)}, \mathbf{X} - \mathbf{X}^{(k)}\rangle + \sum_{i=1}^{N}\frac{\rho}{2}\langle\mathbf{f}_i^{(k)}, \mathbf{x}_i - \mathbf{x}_i^{(k)}\rangle + \frac{1}{\tau}\|\mathbf{X} - \mathbf{X}^{(k)}\|_{\mathrm{F}}^2\Big\}. \tag{15}$$

The optimization in (15) can be decomposed into N separable sub-problems as

$$\mathbf{x}_i^{(k+1)} \approx \min_{\mathbf{x}_i} \Big\{ \alpha_i^{(k)} \|\mathbf{x}_i\|_2 + \frac{1}{2} \langle \mathbf{g}_i^{(k)} + \rho\, \mathbf{f}_i^{(k)}, \mathbf{x}_i - \mathbf{x}_i^{(k)} \rangle + \frac{1}{\tau} \|\mathbf{x}_i - \mathbf{x}_i^{(k)}\|_2^2 \Big\}, \ \forall i \in \mathcal{N}, \tag{16}$$

where $\mathbf{g}_i \in \mathbb{C}^M$ is ith row of $\mathbf{G}$ in (13). By adding an appropriate constant term and with applying some manipulations, we can reformulate the objective function in (16) as

$$\frac{\tau \alpha_i^{(k)}}{2} \|\mathbf{x}_i\|_2 + \frac{1}{2} \|\mathbf{x}_i - \mathbf{c}_i^{(k)}\|_2^2, \quad \forall i \in \mathcal{N}, \tag{17}$$

where $\mathbf{c}_i^{(k)} = \mathbf{x}_i^{(k)} - \frac{\tau}{2}(\mathbf{g}_i^{(k)} + \rho \mathbf{f}_i^{(k)})$. The expression in (17) admits a closed-form solution to (16), given as [19]

$$\mathbf{x}_i^{(k+1)} = \frac{\max\left\{0, \|\mathbf{c}_i^{(k)}\|_2 - \frac{\tau \alpha_i^{(k)}}{2}\right\}}{\|\mathbf{c}_i^{(k)}\|_2} \mathbf{c}_i^{(k)}, \ \forall i \in \mathcal{N}. \tag{18}$$

IV. Numerical Results

Numerical results are provided to show the performance in terms of user activity detection accuracy, channel estimation quality, and convergence rate for the proposed method with comparison to different MMV reconstruction algorithms.

A. Simulation Setup

The BS employs a ULA with $M = 20$ antennas. A total of $N = 200$ users are randomly distributed in a cell with the radius of 100 m. We set the channel angular spread to $\zeta = 5°$ and we fix the number of propagation channel paths between the ith user and the BS to $P_i = 200$, $\forall i \in \mathcal{N}$. At each coherence interval T_c, a fixed number of $K = 10$ users are active. Each user $i \in \mathcal{N}$ is assigned with a unique normalized quadratic phase shift keying (QPSK) sequence $\boldsymbol{\phi}_i$. The QPSK pilot symbols are drawn from an i.i.d. complex Bernoulli distribution. In addition, the BS uses $T = 100$ channel realizations in the training phase to estimate $\{\hat{\mathbf{R}}_i\}_{i=1}^N$.

The performance of the proposed method is compared to the following algorithms that solve the sparse MMV problem in (4): 1) fast alternating direction methods (F-ADM) [14], 2) SPARROW (based on $\ell_{2,1}$-norm minimization) [13], and 3) a least square (LS) estimator that is provided the "oracle" knowledge on the true set of active users, i.e., set $\mathcal{S}$.

The recovery performance is quantified in terms of normalized mean square error (NMSE) and successful support recovery (SR) rate. Let $\mathbf{X}_\mathcal{S}$ and $\hat{\mathbf{X}}_\mathcal{S}$ denote the original and estimated channel matrix, respectively, restricted to the true active support $\mathcal{S}$. Accordingly, NMSE is defined as $\frac{\mathbb{E}[\|\mathbf{X}_\mathcal{S} - \hat{\mathbf{X}}_\mathcal{S}\|_\mathrm{F}^2]}{\mathbb{E}[\|\mathbf{X}_\mathcal{S}\|_\mathrm{F}^2]}$. We present NMSE as normalized average square error (NASE) computed via Monte-Carlo averaging on 10^3 realizations over the randomness of channel $\mathbf{X}$, pilot sequence $\boldsymbol{\Phi}$, and noise $\mathbf{W}$. The SR rate is defined as $\frac{|\mathcal{S} \cap \hat{\mathcal{S}}|}{|\mathcal{S} - \hat{\mathcal{S}}| + K}$, where $\hat{\mathcal{S}} = \{i \,|\, \|\mathbf{x}_i\|_2 > \delta, \forall i \in \mathcal{N}\}$ denotes the detected support for a pre-defined threshold parameter δ. Thus, $|\mathcal{S} \cap \hat{\mathcal{S}}|$ denotes the number of correctly identified active users, whereas $|\mathcal{S} - \hat{\mathcal{S}}|$ accounts both for the number of misdetected active users and falsely identified inactive users.

The recovery performance of the algorithms depends on their hyper-parameters. For all algorithms, the sparsity of the solution is controlled by β_1 which is usually selected via a trial and error approach. In our simulations, we set β_1 based on the noise variance σ^2 as $\beta_1 = \sqrt{\frac{\sigma^2}{2}}$ since it provided the most robust convergence for the SNR range $[0 - 20]$ dB. The proposed approach depends also on β_2, which controls the emphasis on the covariance deviation term. We used empirical tuning and found a robust choice for β_2 to be 25% of the average norm of the effective channels.

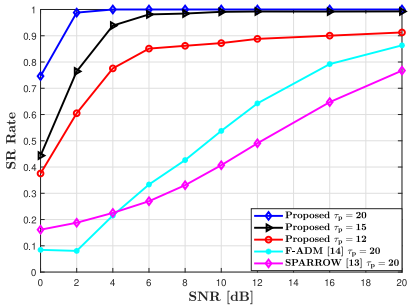

Fig. 2. User activity detection performance in terms of SR rate vs. the SNR.

B. Results

The performance of the different algorithms in terms of the SR rate against SNR is shown in Fig. 2. The results show that for pilot sequence length $\tau_\mathrm{p} = 20$, the proposed method indisputably provides the highest SR rate amongst all the considered algorithms. The proposed method identifies the set of true active users $\mathcal{S}$ perfectly for SNR > 10 dB. Furthermore, reducing the pilot length by a factor of 25 % (i.e., $\tau_\mathrm{p} = 15$) affects the performance of the proposed method only moderately. Hence, the significant reduction in signalling overhead clearly outweighs the minor performance degradation. More interestingly, the results indicate that even with 40% reduction in the pilot sequence length (i.e., $\tau_\mathrm{p} = 12$), the proposed approach outperforms both ADM and SPARROW that use $\tau_\mathrm{p} = 20$. The obtained results imply that utilizing the prior channel covariance information is highly beneficial in the identification of the true set of active users.

Now, we focus on the performance of channel estimation. Fig. 3 compares the NASE performance for the different algorithms. It can be seen that for the pilot sequence length of $\tau_\mathrm{p} = 20$, the proposed method provides the best performance. The proposed method achieves the same performance as the oracle LS with 4 dB lower SNR. Furthermore, Fig. 3 illustrates one advantageous feature of utilizing prior channel statistics information: for SNR < 16 dB, the proposed method with $\tau_\mathrm{p} = 15$ achieves the same performance as oracle LS that uses a pilot length of $\tau_\mathrm{p} = 20$. Moreover, even when using pilot sequence length of $\tau_\mathrm{p} = 12$, the proposed method achieves better performance compared to ADM and SPARROW.

The results in Figs. 2 and 3 show that the performance of the proposed approach depends heavily on the length of the pilot sequence. Although user activity detection achieves near-optimal performance at $\tau_\mathrm{p} = 1.5K$, channel estimation accuracy keeps significantly improving with the increase of τ_p. Nevertheless, increasing τ_p will come at the cost of large signalling overhead. Therefore, a trade-off has to be made between recovery performance and signalling overhead in the design of the optimal pilot sequence length.

Fig. 4 shows the typical convergence behavior of the proposed algorithm at SNR $= 20$ dB. The results reveal that for $\tau_\mathrm{p} = 20$, both the proposed algorithm and F-ADM require approximately 20 iterations to convergence. However,

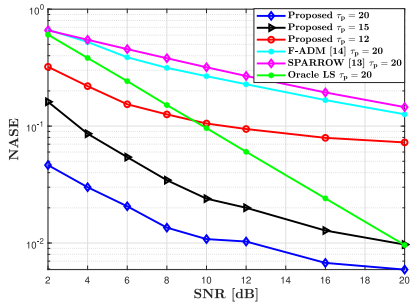

Fig. 3. Channel estimation performance in terms of NMSE vs. the SNR.

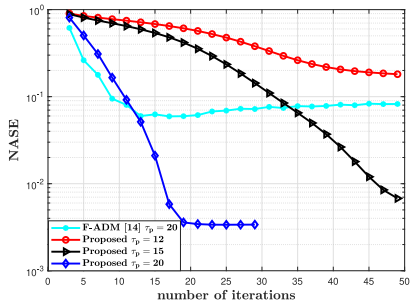

Fig. 4. NASE vs. the number of iterations at SNR $= 20$ dB.

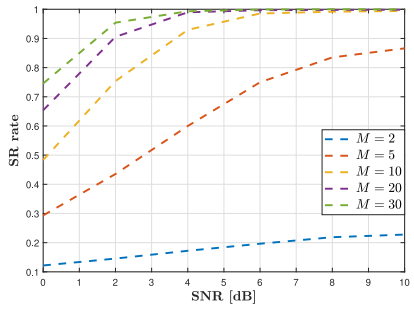

Fig. 5. User activity detection performance vs. the number of BS antennas.

the performance of the proposed approach is significantly better in terms of NASE. Furthermore, reducing the pilot length results in a slower convergence rate: for instance, for $\tau_{\mathrm{p}} = 15$, the proposed approach requires up to 50 iterations to converge.

Fig. 5 shows the active user detection performance as a function of SNR for different numbers of BS antennas M. The results show that increasing the number of antennas from $M = \frac{K}{2}$ to $M = K$ provides more gains than the increase from $M = K$ to $M = 3K$; this means that the relative improvement in user activity detection gradually decreases as M increases. This result indicates that in the massive MIMO regime ($M > K$), the increase in the pilot length provides more significant gains than the increase in the number of antennas M.

V. Conclusion

This letter addressed the JUICE problem in a spatially correlated fading channel for mMTC. We developed a novel JUICE problem formulation that exploits prior information about second order channel statistics. An ADMM-based algorithm was derived to provide computationally efficient solution. The numerical results show significant improvement in the user activity detection accuracy and channel estimation performance obtained by the proposed method.

References

[1] C. Bockelmann *et al.*, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[2] A. C. Cirik, N. M. Balasubramanya, L. Lampe, G. Vos, and S. Bennett, "Toward the standardization of grant-free operation and the associated NOMA strategies in 3GPP," *IEEE Commun. Stand. Mag.*, vol. 3, no. 4, pp. 60–66, Dec. 2019.

[3] E. J. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] 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.

[5] 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.

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

[7] Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Covariance based joint activity and data detection for massive random access with massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019, pp. 1–6.

[8] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. 50th Asilomar Conf. Signals Syst. Comput.*, 2016, pp. 974–978.

[9] B. Clerckx, G. Kim, and S. Kim, "Correlated fading in broadcast MIMO channels: Curse or blessing?" in *Proc. IEEE GLOBECOM*, 2008, pp. 1–5.

[10] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[11] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—-What is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, pp. 3–20, Nov. 2019.

[12] L. Sanguinetti, E. Björnson, and J. Hoydis, "Towards massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, Jan. 2020.

[13] C. Steffens, M. Pesavento, and M. E. Pfetsch, "A compact formulation for the $l_{2,1}$ mixed-norm minimization problem," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1483–1497, Mar. 2018.

[14] H. Lu, X. Long, and J. Lv, "A fast algorithm for recovery of jointly sparse vectors based on the alternating direction methods," in *Proc. 14th Int. Conf. Artif. Intell. Stat. (AISTATS)*, 2011, pp. 461–469.

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

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[17] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," in *Wavelets and Sparsity XV*, vol. 8858. San Diego, CA, USA: Int. Soc. Opt. Photon., 2013, Art. no. 88580R. doi: 10.1117/12.2024410

[18] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing," *SIAM J. Imag. Sci.*, vol. 1, no. 1, pp. 143–168, 2008.

[19] T. Goldstein, C. Studer, and R. Baraniuk, "A field guide to forward-backward splitting with a FASTA implementation," 2014. [Online]. Available: arXiv:1411.3406.