Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Joint active user detection and channel estimation for massive machine-type communications: a difference-of-convex optimization perspective*

Lijun ZHU[1], Kaihui LIU[†‡2], Liangtian WAN[3], Lu SUN[4], Yifeng XIONG[1]

[1]*School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China*

[2]*School of Electrical Engineering & Intelligentization, Dongguan University of Technology, Dongguan 523808, China*

[3]*School of Software, Dalian University of Technology, Dalian 116024, China*

[4]*Information Science Technology College, Dalian Maritime University, Dalian 116026, China*

[†]E-mail: kaihuiL@outlook.com

Received Jan. 15, 2024; Revision accepted Apr. 9, 2024; Crosschecked Mar. 12, 2025

Abstract: Sparsity-based joint active user detection and channel estimation (JADCE) algorithms are crucial in grant-free massive machine-type communication (mMTC) systems. The conventional compressed sensing algorithms are tailored for noncoherent communication systems, where the correlation between any two measurements is as minimal as possible. However, existing sparsity-based JADCE approaches may not achieve optimal performance in strongly coherent systems, especially with a small number of pilot subcarriers. To tackle this challenge, we formulate JADCE as a joint sparse signal recovery problem, leveraging the block-type row-sparse structure of millimeter-wave (mmWave) channels in massive multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems. Then, we propose an efficient difference-of-convex function algorithm (DCA) based JADCE algorithm with multiple measurement vector (MMV) frameworks, promoting the row-sparsity of the channel matrix. To mitigate the computational complexity further, we introduce a fast DCA-based JADCE algorithm via a proximal operator, which allows a low-complexity alternating direction multiplier method (ADMM) to resolve the optimization problem directly. Finally, simulation results demonstrate that the two proposed difference-of-convex (DC) algorithms achieve effective active user detection and accurate channel estimation compared with state-of-the-art compressed sensing based JADCE techniques.

Key words: Joint active user detection and channel estimation; Massive machine-type communications; Difference-of-convex function algorithm; Alternating direction multiplier method

https://doi.org/10.1631/FITEE.2400035 **CLC number:** TP301.6

1 Introduction

Millimeter-wave (mmWave) massive multiple-input multiple-output (massive MIMO) systems are remarkable for their gigabit-per-second high-speed data transmission rates and are considered a critical technology that can advance the sixth generation (6G) wireless communication networks. To mitigate severe propagation loss, numerous antenna elements are required by the base station (BS) for

‡ Corresponding author

* Project supported by the Guangdong Basic and Applied Basic Research Foundation, China (No. 2022A1515140074) and the Natural Science Foundation of Liaoning Province, China (No. 2023-MS-108)

ORCID: Lijun ZHU, https://orcid.org/0009-0002-7363-3354; Kaihui LIU, https://orcid.org/0000-0001-8885-6767; Liangtian WAN, https://orcid.org/0000-0003-0574-8360; Lu SUN, https://orcid.org/0000-0001-7779-4484; Yifeng XIONG, https://orcid.org/0000-0002-4290-7116

beamforming. However, the conventional antennas are associated with considerable hardware costs and power consumption owing to the large number of radio frequency (RF) chains required (Chukhno et al., 2024). To address this problem, a massive multipanel MIMO system integrates the antenna elements into a uniform antenna array with a partially connected hybrid structure, thus demonstrating advantages such as spatial diversity, flexible array deployment, and reduced hardware cost and power consumption. As a result, the massive multipanel MIMO systems have emerged as ideal array configurations for mmWave communications.

In addition, massive machine-type communication (mMTC) is an attractive application scenario for the 6G wireless communication networks for supporting various Internet of Things (IoT) applications. Compared to traditional communication schemes, mMTC is characterized by large-scale user connections, short packet transfers, low power consumption, and sporadic communication (i.e., only a fraction of users are active at any given coherent time interval). In view of avoiding the high signal overhead and high delay of traditional grant-based random access solutions, the grant-free random access protocol is popularly considered a candidate technology for the 6G wireless networks (Gao et al., 2024). Using this protocol, active users can transmit data signals by delivering preallocated pilot sequences without authorization from the BS. On this basis, it is paramount to perform joint active user detection and channel estimation (JADCE), which detects active users for ensuring efficient utilization of spectrum resources and obtains accurate channel state information (CSI) for improving communication quality and reliability (Liu KH et al., 2023).

Due to the sparseness of the user activity patterns, the JADCE problem can be expressed as a joint sparse signal recovery problem, which can be resolved by various compressed sensing algorithms (Liu KH et al., 2019, 2022, 2023; Gan et al., 2021; Wan et al., 2022; Zhu et al., 2023). Based on the grant-free nonorthogonal multi-access systems, Li S et al. (2021) and Guo YR et al. (2024) proposed lower computational complexity orthogonal matching pursuit (OMP) and subspace (SP) algorithms for JADCE, respectively. Wang BC et al. (2016) used compressed sampling matching pursuit (CoSaMP) to detect user activities within multiple time slots by identifying the nonzero element locations of sparse signals. However, these greedy sparse signal reconstruction algorithms (Wang BC et al., 2016; Li S et al., 2021; Guo YR et al., 2024) cannot leverage effectively any prior information; moreover, the inverse of the high-dimensional matrix increases the computational complexity due to the large number of users.

To mitigate the computational complexity, Liu L and Yu (2018) introduced an approximate message passing (AMP) based scheme for massive connectivity, achieving perfect active device detection but encountering channel estimation errors due to the numerous BS antennas. Zhang XX et al. (2021) proposed two sparse Bayesian learning (SBL) approaches, including SBL with Gaussian prior information and fast inverse-free SBL (FI-SBL), to enhance active user detection robustness and channel estimation accuracy. Subsequently, incorporating the generalized AMP (GAMP) algorithm into SBL and the pattern-coupled SBL (PC-SBL) algorithms, Zhang XX et al. (2023) introduced two Bayesian algorithms, GAMP-SBL and GAMP-PCSBL, aiming to mitigate computational complexity in JADCE problems. Additionally, variational Bayesian inference (VBI) was adopted in Zhang YY et al. (2018) and Zhang ZJ et al. (2019) for active user detection and channel estimation leveraging mean-field AMP and Gaussian AMP, respectively. Thereafter, Zhang ZJ et al. (2023) incorporated the AMP algorithm with VBI-based clustering, and presented the AMP-combined VBI clustering (AMP-VBIC) algorithm to tackle the problem of joint active user detection and data detection. By unfolding the AMP method into a forward deep neural network, Cui et al. (2021) and Zheng et al. (2024) proposed new model-driven deep learning algorithms for large-scale connectivity scenarios. However, these approaches (Wang BC et al., 2016; Liu L and Yu, 2018; Zhang YY et al., 2018; Zhang ZJ et al., 2019; Zhang XX et al., 2021, 2023; Ma et al., 2024) overlook massive MIMO systems.

Based on the spatially correlated massive MIMO channels, Djelouat et al. (2022) proposed a mixed norm minimization formulation by leveraging prior knowledge of channel distribution, thus enhancing channel estimation quality and user activity detection accuracy. By exploiting auxiliary information for channel decoding and common sparsity of the received data signals, Bian et al. (2023, 2024)

introduced the bilinear generalized AMP (BiG-AMP) and free probability AMP (FPAMP) algorithms for joint active device detection and channel estimation, respectively. The efficient message passing (MP) algorithm was proposed by Wei et al. (2022) to address channel estimation and signal recovery challenges in reconfigurable intelligent surfaces (RISs) assisted wireless communication systems. Rajoriya and Budhiraja (2023) suggested a Bayesian scheme for incorporating AMP into SBL, offering excellent SBL performance with low complexity of AMP. Li Y et al. (2024) introduced a task-driven activity (TDA) model and proposed an MP-based parameter estimation algorithm for joint device activity detection and channel estimation. For multicarrier massive MIMO orthogonal frequency division multiplexing (OFDM) systems, Ke et al. (2020) proposed a generalized multiple measurement vector AMP (GMMV-AMP) algorithm leveraging channel sparsity in spatial and angular domains for JADCE. When the sensing matrix was a partial unitary matrix, Xiu et al. (2023) used an expectation maximization algorithm to adaptively learn prior parameters and proposed the orthogonal AMP (OAMP) algorithm for the mmWave massive multipanel MIMO systems. Subsequently, Mei et al. (2023) proposed a two-stage OAMP method for the JADCE problem over near-field channels. Guo MQ and Gursoy (2023) used the GAMP algorithm to address equivalent channel estimation and optimal fusion rule based active device detection in centralized and distributed RISs-assisted large-scale connectivity scenarios. By leveraging pilot sequence structures, Marata et al. (2023) proposed the approximation error method (AEM) alternating direction method of multipliers (AEM-ADMM) and AEM-SBL algorithms for JADCE. Additionally, a new training sequence padding (TSP) multicarrier system (TSP-MCS) was proposed in Ying et al. (2023) for low Earth orbit (LEO) satellite constellations, in which the training sequence (TS) was designed to enforce JADCE and multipath interference cancellation (MIC). However, these AMP-type algorithms rely on the channel's prior distribution and noise variance, which are challenging to obtain in practice.

In the mMTC scenarios, many problems are coherent. For example, the coherence between the users and BS is related to the distance between them, and the BS and the randomly distributed users typically present strong coherence within a cell; i.e., the coherence is strong when the user is located close to the BS, and conversely, the coherence is weak when the distance between the users and BS is large. However, conventional greedy and Bayesian algorithms are suitable for the noncoherent situation. In the highly coherent scenarios, these algorithms make it difficult to distinguish the channel components efficiently and fail to provide the sparsest solution for the channel vectors. To address this challenge, Yin et al. (2015) proposed a non-convex and Lipschitz continuous metric based ℓ_{1-2} minimization algorithm for single measurement vector (SMV) models, and this algorithm has been widely applied in several fields, including magnetic resonance imaging (MRI), phantom image restoration, seismic attenuation compensation, and channel estimation (Yu et al., 2020). Inspired by the success of the ℓ_{1-2} minimization algorithm in the SMV model, we present the difference-of-convex function algorithm (DCA) and fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithms for enhancing sparsity in MMV models. These algorithms enhance the sparsity of the channel matrices without requiring prior channel distribution information, by using Chorisky decomposition to avoid matrix inverse operations.

The main contributions of this article can be generalized as follows:

1. By exploiting the block-type row-sparse structure of the mmWave channel matrix, we formulate the JADCE problem as a joint sparse signal recovery problem, where compressed sensing algorithms are feasible schemes.

2. Conventional compressed sensing methods may not work well in strongly coherent scenarios, which can suffer from performance loss as the number of pilot subcarriers decreases. As a result, we propose the DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm for the JADCE problem.

3. To solve the DCA efficiently, the convex subproblems decomposed by the original DCA are resolved by the ADMM solver.

4. To further reduce the computational complexity, a fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm is then introduced to solve the JADCE problem. Specifically, we deduce the closed-form solution of the $\mathcal{L}_{2,1-\mathrm{F}}$ metric-proximal operator, which permits a low-complexity ADMM solver to resolve the minimization algorithm in a straightforward manner.

In this article, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively; a, $\boldsymbol{a}$, and $\boldsymbol{A}$ denote scaler, column vector, and matrix, respectively; $\boldsymbol{A}^{\mathrm{T}}$, $\boldsymbol{A}^{\mathrm{H}}$, and $\boldsymbol{A}^{-1}$ indicate the transpose, conjugate transpose, and matrix inverse operations of matrix $\boldsymbol{A}$, respectively; a_m denotes the m^{th} term of vector $\boldsymbol{a}$, and $A_{m,n}$ refers to the element of the m^{th} row and n^{th} column of matrix $\boldsymbol{A}$. Then, $\|\boldsymbol{A}\|_2$, $\|\boldsymbol{A}\|_{2,0}$, $\|\boldsymbol{A}\|_{2,1}$, and $\|\boldsymbol{A}\|_{\mathrm{F}}$ stand for the (2)-norm, $(2,0)$-norm, $(2,1)$-norm, and F-norm of matrix $\boldsymbol{A}$, respectively. $\mathcal{U}(\boldsymbol{A})$ denotes the optimization function with respect to matrix $\boldsymbol{A}$, and $\partial(\mathcal{U}(\boldsymbol{A}))$ is the subgradient of the function of matrix $\boldsymbol{A}$; $\arg\min_{\boldsymbol{A}}\mathcal{U}(\boldsymbol{A})$ denotes the value of variable $\boldsymbol{A}$ that minimizes the function $\mathcal{U}(\boldsymbol{A})$ with respect to matrix $\boldsymbol{A}$. The zero matrix and identity matrix of order N are denoted as $\boldsymbol{0}_N$ and $\boldsymbol{I}_N$, respectively. $\langle \boldsymbol{A}, \boldsymbol{B}\rangle$ is the inner product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and $\otimes$ indicates the Kronecker product operation; $\mathrm{vec}(\boldsymbol{A})$ represents the matrix vectorization of matrix $\boldsymbol{A}$. $\mu(\boldsymbol{A})$ denotes the coherence coefficient of matrix $\boldsymbol{A}$. Finally, $\mathcal{CN}\left(0,\sigma^2\right)$ indicates a complex Gaussian distribution with mean 0 and covariance σ^2.

2 System model

In this section, we consider a hybrid precoding-based mmWave massive MIMO-OFDM uplink system and formulate the JADCE problem.

2.1 Channel model

In grant-free mMTC systems, the BS employs a partially connected hybrid MIMO-OFDM system with multipanel array antennas (Wang W et al., 2018), and communicates with K randomly distributed potential single-antenna users. The configuration of the multipanel array antennas is presented in Fig. 1. The BS is equipped with $N_{\mathrm{p}} = I_{\mathrm{h}}I_{\mathrm{v}}$ antenna array panels, each of which uses a uniform planar array (UPA). Here, I_{h} and I_{v} represent the numbers of subarray panels in the horizontal and vertical dimensions of the multipanel antenna array, respectively. Each subarray panel comprises $M_{\mathrm{BS}} = M_{\mathrm{h}}M_{\mathrm{v}}$ antennas, where M_{h} and M_{v} denote the numbers of antennas in the horizontal and vertical dimensions of each subarray panel, respectively. Consequently, the numbers of antennas in the horizontal and vertical directions of the multipanel array are $N_{\mathrm{h}} = I_{\mathrm{h}}M_{\mathrm{h}}$ and $N_{\mathrm{v}} = I_{\mathrm{v}}M_{\mathrm{v}}$, respectively, and the total number of BS antennas equipped is $N_{\mathrm{BS}} = N_{\mathrm{h}}N_{\mathrm{v}}$. The BS leverages N_{P} radio frequency chains, each of which is connected to the corresponding subarray panel via a partially connected phase-shift network. The adjacent antenna spacing for each subarray panel is $d = \lambda/2$, where λ is the carrier wavelength, and the adjacent panel spacing is $\Delta = Dd$, where D is an integer with $D \geq 2$ (Xiu et al., 2023).

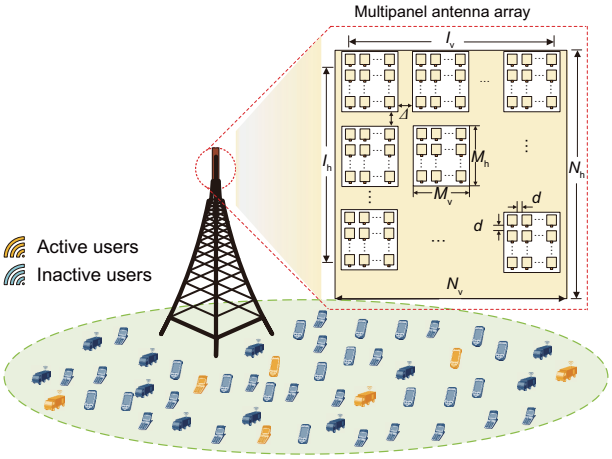

Fig. 1 In the mMTC scenarios, the massive multipanel MIMO systems employ the mmWave channel model

To mitigate the frequency-selective fading caused by the wireless channel multipath effect, the OFDM technology with N_{c} subcarriers is employed to access the substantial IoT. The P pilot subcarriers are distributed uniformly across the N_{c} subcarriers to transmit pilot signals for the JADCE problem (Alkhateeb et al., 2014). Since the mmWave channel experiences limited scattering paths, we exploit a geometric wideband mmWave channel model along with L scattering paths for a massive multipanel MIMO-OFDM system. Each scattering path corresponds to one path delay and one angle of arrival (AoA), leading to inherent angular domain sparsity in a limited scattering environment. As a result, the p^{th} pilot subcarrier in the t^{th} OFDM system of the mmWave channel $\boldsymbol{h}_{p,k}^{t} \in \mathbb{C}^{N_{\mathrm{BS}}}$ between the BS and the k^{th} user can be represented by

$$\begin{aligned}&\boldsymbol{h}_{p,k}^{t}\\&=\sum_{l=1}^{L}\beta_{k,l}^{t}\boldsymbol{a}_{\mathrm{MP}}\left(\mu_{k,l}^{t},\psi_{k,l}^{t}\right)\mathrm{e}^{-\mathrm{j}2\pi\tau_{k,l}^{t}\left(-\frac{B_{\mathrm{s}}}{2}+\left(\frac{pN_{\mathrm{c}}}{P}-1\right)\frac{B_{\mathrm{s}}}{N_{\mathrm{c}}}\right)},\end{aligned}\tag{1}$$

where B_{s} denotes the bilateral signal bandwidth, $\beta_{k,l}^t$ and $\tau_{k,l}^t$ are the complex gain and time delay of the l^{th} path for the k^{th} user, respectively, and we assume that the path amplitudes follow a Gaussian distribution, i.e., $\beta_{k,l}^t \sim \mathcal{CN}(0,1)$. Then $\boldsymbol{a}_{\mathrm{MP}}\left(\mu_{k,l}^t, \psi_{k,l}^t\right)$ represents the received array steering vector which can be calculated explicitly by $\boldsymbol{a}_{\mathrm{MP}}\left(\mu_{k,l}^t, \psi_{k,l}^t\right) = \mathrm{vec}\left[\boldsymbol{A}\left(\mu_{k,l}^t, \psi_{k,l}^t\right)\right]$, where $\mu_{k,l}^t = \pi \sin\theta_{k,l}^t \cos\phi_{k,l}^t$ and $\psi_{k,l}^t = \pi \sin\phi_{k,l}^t$ denote the horizontal and vertical virtual AoAs, respectively, and $\theta_{k,l}^t$ and $\phi_{k,l}^t$ denote the azimuth and elevation AoAs, respectively. In the following, we calculate $\boldsymbol{A}\left(\mu_{k,l}^t, \psi_{k,l}^t\right) = \boldsymbol{a}_{\mathrm{h}}\left(\mu_{k,l}^t\right)\boldsymbol{a}_{\mathrm{v}}^{\mathrm{T}}\left(\psi_{k,l}^t\right)$ (Wang W et al., 2018). The array steering vector in the horizontal direction is given by $\boldsymbol{A}_{\mathrm{h}}\left(\mu_{k,l}^t\right) = \boldsymbol{A}_{I_{\mathrm{h}}}\left(\mu_{k,l}^t\right) \otimes \boldsymbol{A}_{M_{\mathrm{h}}}\left(\mu_{k,l}^t\right) \in \mathbb{C}^{N_{\mathrm{h}}}$, in which

$$\begin{aligned}&\boldsymbol{a}_{I_{\mathrm{h}}}\left(\mu_{k,l}^t\right)\\ &= \left[1, \mathrm{e}^{\mathrm{j}(M_{\mathrm{h}}+D-1)\mu_{k,l}^t}, \cdots, \mathrm{e}^{\mathrm{j}(I_{\mathrm{h}}-1)(M_{\mathrm{h}}+D-1)\mu_{k,l}^t}\right]^{\mathrm{T}},\end{aligned} \tag{2}$$

$$\boldsymbol{a}_{M_{\mathrm{h}}}\left(\mu_{k,l}^t\right) = \left[1, \mathrm{e}^{\mathrm{j}\mu_{k,l}^t}, \cdots, \mathrm{e}^{\mathrm{j}(M_{\mathrm{h}}-1)\mu_{k,l}^t}\right]^{\mathrm{T}}, \tag{3}$$

and the horizontal array response vectors are associated with the horizontal and vertical multipanel arrays. Similarly, the array steering vector in the vertical direction is given by $\boldsymbol{a}_{\mathrm{v}}\left(\psi_{k,l}^t\right) = \boldsymbol{a}_{I_{\mathrm{v}}}\left(\psi_{k,l}^t\right) \otimes \boldsymbol{a}_{M_{\mathrm{v}}}\left(\psi_{k,l}^t\right) \in \mathbb{C}^{N_{\mathrm{v}}}$, where vertical array response vectors associated with the horizontal and vertical multipanel arrays $\boldsymbol{A}_{I_{\mathrm{v}}}\left(\psi_{k,l}^t\right) \in \mathbb{C}^{I_{\mathrm{v}}}$ and $\boldsymbol{A}_{M_{\mathrm{v}}}\left(\psi_{k,l}^t\right) \in \mathbb{C}^{M_{\mathrm{v}}}$ are written respectively by substituting $\mu_{k,l}^t$ and I_{h} with $\psi_{k,l}^t$ and I_{v} in Eq. (2) along with substituting $\mu_{k,l}^t$ and M_{h} with $\psi_{k,l}^t$ and M_{v} in Eq. (3), respectively.

2.2 Hybrid precoding-based massive MIMO-OFDM systems

In the uplink grant-free mMTC scenarios, owing to the sporadic traffic of the number of IoT users K, only a tiny fraction of active users, denoted by $K_\alpha\,(K_\alpha \ll K)$, transmit pilot sequences to the BS during any given time interval, assuming that CSI is slowly changing during this period (Qiao et al., 2022). The activity of each user is mutually independent: when user k is active, $\alpha_k = 1$; otherwise, $\alpha_k = 0$. We contemplate a hybrid precoding-based mmWave massive multipanel MIMO-OFDM system, and the received pilot signal $\boldsymbol{y}_p^t \in \mathbb{C}^{N_{\mathrm{BS}}}$ from the t^{th} OFDM symbol with the p^{th} pilot subcarrier is expressed as

$$\begin{aligned}\boldsymbol{y}_p^t &= \left(\boldsymbol{W}_{\mathrm{RF}}^t \boldsymbol{W}_{\mathrm{BB}}\right)^{\mathrm{H}} \sum_{k=1}^{K} \alpha_k \boldsymbol{h}_{p,k}^t s_{p,k}^t + \left(\boldsymbol{W}_{\mathrm{RF}}^t \boldsymbol{W}_{\mathrm{BB}}\right)^{\mathrm{H}} \overline{\boldsymbol{n}}_p^t\\ &= \left(\boldsymbol{W}_{\mathrm{RF}}^t \boldsymbol{W}_{\mathrm{BB}}\right)^{\mathrm{H}} \widetilde{\boldsymbol{H}}_p^t \boldsymbol{s}_p^t + \boldsymbol{n}_p^t,\end{aligned} \tag{4}$$

where $\widetilde{\boldsymbol{H}}_p^t = \left[\alpha_1 \boldsymbol{h}_{p,1}^t, \alpha_2 \boldsymbol{h}_{p,2}^t, \cdots, \alpha_K \boldsymbol{h}_{p,K}^t\right] \in \mathbb{C}^{N_{\mathrm{BS}} \times K}$ is the composite channel matrix, which contains the binary activity indicator flag of users and the information of the channel, appealing us to estimate the channel matrix and detect the active users simultaneously. Then $\boldsymbol{s}_p^t = \left[s_{p,1}^t, s_{p,2}^t, \cdots, s_{p,K}^t\right]^{\mathrm{T}} \in \mathbb{C}^K$ is a randomly selected pilot signal vector from the columns of $\boldsymbol{D}_K$, and $\boldsymbol{D}_K$ is represented by a $K \times K$ Gaussian matrix. The composite additive noise vector is $\boldsymbol{n}_p^t = \left(\boldsymbol{W}_{\mathrm{RF}}^t \boldsymbol{W}_{\mathrm{BB}}\right)^{\mathrm{H}} \overline{\boldsymbol{n}}_p^t$, and $\overline{\boldsymbol{n}}_p^t \in \mathbb{C}^{N_{\mathrm{BS}}}$ is the additive Gaussian white noise vector that corrupts the received signal, i.e., $\bar{\boldsymbol{n}}_p^t \sim \mathcal{CN}\left(0, \sigma^2 \boldsymbol{I}_{N_{\mathrm{BS}}}\right)$ with the variance of σ^2 (Tong et al., 2021).

Owing to the sporadic communication pattern of the active users in massive connectivity, all antennas have identical sparsity in the spatial-domain structure (Ke et al., 2020), which indicates that the rows of $\widetilde{\boldsymbol{H}}_p^t$ corresponding to inactive users are zero; thus, $\widetilde{\boldsymbol{H}}_p^t$ has a sparse structure with K_α non-zero rows. The sparsity of active users is identical within all subchannels, and $\left\{\widetilde{\boldsymbol{H}}_p^t\right\}_{p=1}^P$ in the frequency domain displays a common sparse pattern, i.e., with the same sparse support set. In conjunction with the above analysis of the channel structure, $\left\{\widetilde{\boldsymbol{H}}_p^t\right\}_{p=1}^P$ possesses sparsity in the spatial-frequency-domain structure given in Fig. 2a.

At the BS side, we employ a partially connected multipanel array architecture (Xiu et al., 2023). $\boldsymbol{W}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{P}} \times N_{\mathrm{P}}}$ is the digital combination matrix, which is defined as $\boldsymbol{W}_{\mathrm{BB}} = \boldsymbol{I}_{N_{\mathrm{P}}}$, and $\boldsymbol{W}_{\mathrm{RF}}^t \in \mathbb{C}^{N_{\mathrm{BS}} \times N_{\mathrm{P}}}$ is the analog combination matrix. The settings are as follows: First, $\boldsymbol{w}_{n_p}^t$ is the n_p^{th} column of $\boldsymbol{W}_{\mathrm{RF}}^t$, which can be represented as $\left[\boldsymbol{w}_{n_p}^t\right]_{\mathcal{I}_{n_p}} = \frac{1}{\sqrt{M_{\mathrm{BS}}}}\left[\boldsymbol{z}_{n_p}^t\right]_{\mathcal{I}_{n_p}}$, and the set of sequences $\mathcal{I}_{n_p}$ denotes the antenna index of the n_p^{th} subarray panel, initialized as $\boldsymbol{w}_{n_p} = \boldsymbol{0}_{N_{\mathrm{BS}}}$. Next, $\boldsymbol{z}_{n_p}^t$ is the n_p^{th} column of $\boldsymbol{Z}$ which is a local Gaussian matrix; i.e., $\boldsymbol{Z} = \boldsymbol{D}_{N_{\mathrm{v}}} \otimes \boldsymbol{D}_{N_{\mathrm{h}}} \boldsymbol{P} \in \mathbb{C}^{N_{\mathrm{BS}} \times N_{\mathrm{P}}}$, the modulus of all

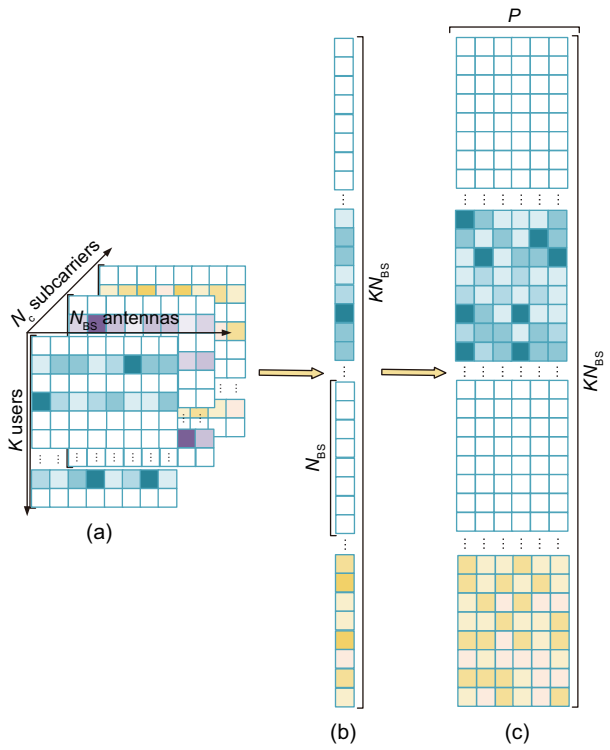

Fig. 2 Due to the spatial-frequency-domain structure, $\widetilde{\boldsymbol{H}}_{\boldsymbol{p}}$ exhibits a row-sparse structure in (a); by vectorization of $\widetilde{\boldsymbol{H}}_{\boldsymbol{p}}$, $\boldsymbol{h}_{\boldsymbol{p}}^{\boldsymbol{t}}$ shows a block-type sparse structure in (b); after aggregating P pilot subcarriers, $\widetilde{\boldsymbol{H}}$ has the property of block-type row sparsity in (c)

elements in $\boldsymbol{Z}$ is 1, and $\boldsymbol{P}$ represents a permutation matrix.

2.3 Problem formulation

We exploit equation $\mathrm{vec}(\boldsymbol{BCD}) = (\boldsymbol{D}^{\mathrm{T}} \otimes \boldsymbol{B})\mathrm{vec}(\boldsymbol{C})$ (Petersen and Pedersen, 2012) to represent the received signal $\boldsymbol{y}_p^t$ as a vector form

$$\boldsymbol{y}_p^t = \widetilde{\boldsymbol{\Phi}}_p^t \boldsymbol{h}_p^t + \boldsymbol{n}_p^t, \tag{5}$$

where $\widetilde{\boldsymbol{\Phi}}_p^t = \left(\boldsymbol{s}_p^t\right)^{\mathrm{T}} \otimes \left(\boldsymbol{W}_{\mathrm{RF}}^t\right)^{\mathrm{H}} \in \mathbb{C}^{N_{\mathrm{p}} \times J}$, $\boldsymbol{h}_p^t = \mathrm{vec}\left(\widetilde{\boldsymbol{H}}_p^t\right) \in \mathbb{C}^J$, and $J = KN_{\mathrm{BS}}$. Due to the row-sparse structure of $\widetilde{\boldsymbol{H}}_p^t$, the elements of $\boldsymbol{h}_p^t$ are non-zero from $(k-1)N_{\mathrm{BS}}+1$ to kN_{BS} concerning the k^{th} active user, and $\boldsymbol{h}_p^t$ has block-type sparse structure in Fig. 2b, enlightening us to detect active users by exploiting the positions of the non-zero elements of the channel matrix (Ke et al., 2020; Shao et al., 2020). Moreover, we consider the same signal vector across all the pilot subcarriers, i.e., when $1 \le p \le P$, $\boldsymbol{s}_p^t = \boldsymbol{s}^t$ and $\widetilde{\boldsymbol{\Phi}}_p^t = \widetilde{\boldsymbol{\Phi}}^t$. Therefore, the received signal from aggregating P pilot subcarriers at the t^{th} OFDM symbol can be denoted as

$$\widetilde{\boldsymbol{Y}}^t = \widetilde{\boldsymbol{\Phi}}^t \widetilde{\boldsymbol{H}}^t + \widetilde{\boldsymbol{N}}^t, \tag{6}$$

where $\widetilde{\boldsymbol{Y}}^t = [\boldsymbol{y}_1^t, \boldsymbol{y}_2^t, \cdots, \boldsymbol{y}_P^t] \in \mathbb{C}^{N_{\mathrm{p}} \times P}$ is the aggregated received signal matrix, $\widetilde{\boldsymbol{N}}^t = [\boldsymbol{n}_1^t, \boldsymbol{n}_2^t, \cdots, \boldsymbol{n}_P^t] \in \mathbb{C}^{N_{\mathrm{p}} \times P}$ is the aggregated noise matrix, $\widetilde{\boldsymbol{H}}^t = [\boldsymbol{h}_1^t, \boldsymbol{h}_2^t, \cdots, \boldsymbol{h}_P^t] \in \mathbb{C}^{J \times P}$ is the aggregated channel matrix, and $\widetilde{\boldsymbol{H}}^t$ has a block-type row-sparse structure as given in Fig. 2c.

During the T continuous OFDM symbols, the structured sparsity of user activity remains unchanged, and the duration of consecutive OFDM symbols is shorter than the channel coherence time of the channel. Therefore, in the large-scale IoT scenarios, the CSI is slowly changing over T consecutive OFDM symbols, which could be considered unchanged, i.e., $\widetilde{\boldsymbol{H}}^t = \widetilde{\boldsymbol{H}}$. As a result, the stacked received signal matrix $\widetilde{\boldsymbol{Y}}$ in the T consecutive OFDM symbols is represented as follows:

$$\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{\Phi}} \widetilde{\boldsymbol{H}} + \widetilde{\boldsymbol{N}}, \tag{7}$$

where $\widetilde{\boldsymbol{Y}} = [(\widetilde{\boldsymbol{Y}}^1)^{\mathrm{T}}, (\widetilde{\boldsymbol{Y}}^2)^{\mathrm{T}}, \cdots, (\widetilde{\boldsymbol{Y}}^T)^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{Q \times P}$, and $\widetilde{\boldsymbol{\Phi}} = [(\widetilde{\boldsymbol{\Phi}}^1)^{\mathrm{T}}, (\widetilde{\boldsymbol{\Phi}}^2)^{\mathrm{T}}, \cdots, (\widetilde{\boldsymbol{\Phi}}^T)^{\mathrm{T}}] \in \mathbb{C}^{Q \times J}$ represents the sensing matrix, $Q < J$, $Q = TN_{\mathrm{p}}$, and $\widetilde{\boldsymbol{N}}$ is the stacked noise matrix.

Taking into account the real-valued counterpart $\boldsymbol{H} \in \mathbb{R}^{2J \times P}$ of the channel matrix $\widetilde{\boldsymbol{H}}$, the linear measurements in the real domain (Liu KH et al., 2022) are given by

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{\Phi}\boldsymbol{H} + \boldsymbol{N} \\ &= \begin{bmatrix} \Re(\widetilde{\boldsymbol{\Phi}}) & -\Im(\widetilde{\boldsymbol{\Phi}}) \\ \Im(\widetilde{\boldsymbol{\Phi}}) & \Re(\widetilde{\boldsymbol{\Phi}}) \end{bmatrix} \begin{bmatrix} \Re(\widetilde{\boldsymbol{H}}) \\ \Im(\widetilde{\boldsymbol{H}}) \end{bmatrix} + \begin{bmatrix} \Re(\widetilde{\boldsymbol{N}}) \\ \Im(\widetilde{\boldsymbol{N}}) \end{bmatrix}, \end{aligned} \tag{8}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{2Q \times 2J}$ and $\boldsymbol{Y} \in \mathbb{R}^{2Q \times P}$ denote the real-valued counterparts of $\widetilde{\boldsymbol{\Phi}}$ and $\widetilde{\boldsymbol{Y}}$, respectively. Since $\boldsymbol{H}$ has a block-like row-sparse structure, the JADCE problem can be formulated as a compressed sensing recovery problem based on the joint sparse MMV framework; then, the JADCE problem can be transformed into the following optimization problem:

$$\min_{\boldsymbol{H} \in \mathbb{R}^{2J \times P}} \left(\frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{H}\|_2^2 + \lambda \|\boldsymbol{H}\|_{2,0}\right), \tag{9}$$

where $\lambda > 0$ is a tuning parameter to counterbalance the data fidelity term $\frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{H}\|_2^2$ with the objective function $\|\boldsymbol{H}\|_{2,0}$. Note that it is different from the parameter settings in Xiu et al. (2023),

and the sensing matrix $\boldsymbol{\Phi}$ exploits a Gaussian matrix in this paper. Currently, compressive sensing algorithms prefer noncoherent communication systems, where the correlation from any two measurements is as small as possible. However, numerous problems are coherent, and the existing sparsity-based JADCE methods do not perform well in the presence of strong coherence, resulting in performance degradation as the number of pilot subcarriers declines. Therefore, we propose an MMV-based non-convex difference-of-convex (DC) optimization algorithm to solve the channel estimation problem and detect the activity of the users (Yu et al., 2020). Moreover, our proposed algorithm can enhance the row sparsity of the channel matrix and achieve accurate sparse channel matrix recovery.

3 Proposed DCA-based JADCE algorithm

In this section, we introduce a modified DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm for the JADCE problem in the MMV case. Moreover, we adopt a low-complexity ADMM solver to address the convex subproblems decomposed by DCA.

3.1 DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm

The design of the measurement matrix $\boldsymbol{\Phi}$ is critical for achieving accurate channel estimation. In traditional compressed sensing algorithms, to reconstruct the channel matrix $\boldsymbol{H}$, the measurement matrix $\boldsymbol{\Phi}$ has a low mutual coherence, which is the maximum normalized inner product of two distinct columns of $\boldsymbol{\Phi}$. However, in practical communication scenarios, where many issues are relevant, the performance of existing compressed perception algorithms degrades as the number of measurements decreases in strongly coherent environments. Here, we define the coherence coefficient of this measurement matrix $\boldsymbol{\Phi}$ (Yin et al., 2015; Lou and Yan, 2018) as follows:

$$\mu(\boldsymbol{\Phi}) = \max_{i \neq j} \frac{\left|\boldsymbol{\Phi}_i^{\mathrm{T}} \boldsymbol{\Phi}_j\right|}{\|\boldsymbol{\Phi}_i\|_2 \|\boldsymbol{\Phi}_j\|_2}, \tag{10}$$

where $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Phi}_j$ denote two arbitrary columns from $\boldsymbol{\Phi}$. When the coherence coefficient of the measurement matrix was high, Lou et al. (2015) demonstrated that the ℓ_{1-2} minimization algorithm strengthened the sparsity of the signal and provided better recovery performance compared to classical compressed sensing algorithms.

In contrast to the previous work that focused on sparse recovery in the SMV case, we consider the problem of joint sparse signal recovery under the MMV framework. To solve the problem, we introduce a novel $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm for the difference of $\|\boldsymbol{H}\|_{2,1}$ and $\|\boldsymbol{H}\|_{\mathrm{F}}$ with the MMV framework as follows:

$$\min_{\boldsymbol{H} \in \mathbb{R}^{2J \times P}} \left(\|\boldsymbol{H}\|_{2,1} - \beta \|\boldsymbol{H}\|_{\mathrm{F}} \right) \text{ s.t. } \boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{H}, \tag{11}$$

where β is an adaptive model parameter, $0 < \beta \leq 1$, and $\|\boldsymbol{H}\|_{2,1}$ and $\|\boldsymbol{H}\|_{\mathrm{F}}$ can enhance the row-sparse structure and the stability of the channel matrix, respectively.

The sparse recovery problem (9) is typically formulated as a minimization problem with an NP-hard ℓ_0 norm. To overcome the computational challenges concerning the ℓ_0 norm, we employ the convex relaxation technique in the form of $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm. By applying this technique, we convert problem (9) into an equivalent computationally tractable problem, which is given by

$$\min_{\boldsymbol{H} \in \mathbb{R}^{2J \times P}} \left(\frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{H}\|_2^2 + \lambda \left(\|\boldsymbol{H}\|_{2,1} - \beta \|\boldsymbol{H}\|_{\mathrm{F}} \right) \right). \tag{12}$$

3.2 DCA algorithm

The objective function in problem (12) is a non-convex optimization problem that can be written as the difference of two convex functions. To solve this problem, we use the DCA which is a nonlinear search algorithm for solving non-convex optimization problems. Specifically, DCA is used to minimize the objective function $\mathcal{U}(\boldsymbol{H}) = \mathcal{M}(\boldsymbol{H}) - \mathcal{W}(\boldsymbol{H})$, which is the difference between two convex functions $\mathcal{M}(\boldsymbol{H})$ and $\mathcal{W}(\boldsymbol{H})$. Therefore, we introduce DCA as the solution to the $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm for the MMV-based JADCE problem.

First, DCA requires the construction of two sequences, namely $\{\boldsymbol{G}^s\}$ and $\{\boldsymbol{H}^s\}$, which represent the candidate sequences of the optimal solution for the primal and dual procedures, respectively (Yin et al., 2015). To implement the DCA solution, the objective function $\mathcal{U}(\boldsymbol{H})$ is computed as follows:

$$\begin{aligned} &\boldsymbol{H}^{s+1} \\ &= \min_{\boldsymbol{H} \in \mathbb{R}^{2J \times P}} \left(\mathcal{M}(\boldsymbol{H}) - \left(\mathcal{W}(\boldsymbol{H}) + \langle \boldsymbol{G}^s, \boldsymbol{H} - \boldsymbol{H}^s \rangle \right) \right), \end{aligned} \tag{13}$$

where $\boldsymbol{G}^s$ is the subgradient of $\mathcal{W}(\boldsymbol{H}^s)$ at $\boldsymbol{H}$, and s is the index of the external iteration. According to the definition of the subgradient, it is shown in Appendix that the objective function $\mathcal{M}(\boldsymbol{H})$ forms a monotonically decreasing sequence, providing that $\mathcal{M}(\boldsymbol{H})$ is bounded from below. As a result, the values of the objective function converge.

Next, we use DCA to address the $\mathcal{L}_{2,1-\mathrm{F}}$ minimization problem. By employing the convex difference function, problem (12) can be transformed into a DC decomposition, which is expressed as follows:

$$\mathcal{U}(\boldsymbol{H})=\left(\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\|_2^2+\lambda\|\boldsymbol{H}\|_{2,1}\right)-\lambda\beta\|\boldsymbol{H}\|_{\mathrm{F}}, \tag{14}$$

where $\mathcal{M}(\boldsymbol{H})=\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\|_2^2+\lambda\|\boldsymbol{H}\|_{2,1}$ and $\mathcal{W}(\boldsymbol{H})=\lambda\beta\|\boldsymbol{H}\|_{\mathrm{F}}$ are two convex subproblems, $\|\boldsymbol{H}\|_{\mathrm{F}}$ is differentiable, and the gradient is given by

$$\boldsymbol{G}^s=\frac{\partial\mathcal{W}(\boldsymbol{H}^s)}{\partial\boldsymbol{H}}=\begin{cases}\boldsymbol{0}, & \text{if } \boldsymbol{H}^s=\boldsymbol{0},\\ -\lambda\beta\frac{\boldsymbol{H}^s}{\|\boldsymbol{H}^s\|_{\mathrm{F}}}, & \text{otherwise.}\end{cases} \tag{15}$$

Therefore, according to the iterative solution formula of DCA, the $\mathcal{L}_{2,1-\mathrm{F}}$ minimization problem in Eq. (14) can be represented as follows:

$$\boldsymbol{H}^{s+1}=\begin{cases}\arg\min\limits_{\boldsymbol{H}}\left(\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\|_2^2+\lambda\|\boldsymbol{H}\|_{2,1}\right), & \text{if } \boldsymbol{H}^s=\boldsymbol{0},\\ \arg\min\limits_{\boldsymbol{H}}\left(\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\|_2^2+\lambda\|\boldsymbol{H}\|_{2,1}\right.\\ \qquad\left.+\left\langle\lambda\beta\frac{\boldsymbol{H}^s}{\|\boldsymbol{H}^s\|_{\mathrm{F}}},\boldsymbol{H}\right\rangle\right), & \text{otherwise.}\end{cases} \tag{16}$$

3.3 ADMM solver for the DCA-based JADCE algorithm

ADMM is a simple but powerful algorithm which is well-suited to distributed optimization. By combining the augmented Lagrangian method and dual decomposition, ADMM can efficiently handle large-scale optimization problems. Therefore, we use the ADMM algorithm (Boyd et al., 2011) to solve the convex subproblems in the $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm. To effectively solve the non-convex minimization problem, we introduce the auxiliary matrix $\boldsymbol{Z}\in\mathbb{R}^{2J\times P}$ to derive the ADMM solver and transform the convex subproblem (16) into the following:

$$\widehat{\boldsymbol{H}}^{s+1}=\arg\min_{\boldsymbol{H}}\left(\frac{1}{2}\left\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\right\|_2^2+\langle\boldsymbol{G}^s,\boldsymbol{H}\rangle+\lambda\|\boldsymbol{Z}\|_{2,1}\right)\text{ s.t. }\boldsymbol{H}-\boldsymbol{Z}=\boldsymbol{0}. \tag{17}$$

The augmented Lagrangian function for Eq. (17) can be expressed as follows:

$$\mathcal{L}_\rho(\widehat{\boldsymbol{H}},\boldsymbol{Z},\boldsymbol{W})=\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\|_2^2+\langle\boldsymbol{G}^s,\boldsymbol{H}\rangle+\lambda\|\boldsymbol{Z}\|_{2,1}+\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{H}-\boldsymbol{Z})+\frac{\rho}{2}\|\boldsymbol{H}-\boldsymbol{Z}\|_{\mathrm{F}}^2, \tag{18}$$

where $\boldsymbol{W}\in\mathbb{R}^{2J\times P}$ is the Lagrangian multiplier and $\rho>0$ is the penalty factor.

ADMM employs the decomposition coordination technique to decompose complex global optimization problems into local subproblems that can be efficiently solved. Then, ADMM minimizes the augmented Lagrangian function by updating the estimated channel matrix $\widehat{\boldsymbol{H}}$, the variable $\boldsymbol{Z}$, and the Lagrangian multiplier $\boldsymbol{W}$ in alternation (Boyd et al., 2011).

1. Update the estimated channel matrix $\widehat{\boldsymbol{H}}$. Minimize $\widehat{\boldsymbol{H}}$ associated with $\mathcal{L}_\rho\left(\widehat{\boldsymbol{H}},\boldsymbol{Z},\boldsymbol{W}\right)$ by fixing $\boldsymbol{Z}$ and $\boldsymbol{W}$, and this is comparable to the following expression:

$$\begin{aligned}\widehat{\boldsymbol{H}}^{r+1}&=\arg\min_{\boldsymbol{H}}\mathcal{L}_\rho\left(\boldsymbol{H},\boldsymbol{Z}^r,\boldsymbol{W}^r\right)\\&=\arg\min_{\boldsymbol{H}}\left(\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\|_2^2+\langle\boldsymbol{G}^r,\boldsymbol{H}\rangle\right.\\&\quad\left.+\left(\boldsymbol{W}^r\right)^{\mathrm{T}}\left(\boldsymbol{H}-\boldsymbol{Z}^r\right)+\frac{\rho}{2}\|\boldsymbol{H}-\boldsymbol{Z}^r\|_{\mathrm{F}}^2\right),\end{aligned} \tag{19}$$

where r is the internal iteration index. The augmented Lagrangian function concerning the estimated channel matrix $\widehat{\boldsymbol{H}}$ is convex and differentiable. Therefore, by setting the derivative of the function with respect to $\widehat{\boldsymbol{H}}$ to zero, it is possible to obtain the optimal solution for $\widehat{\boldsymbol{H}}$, which can be expressed as follows:

$$\widehat{\boldsymbol{H}}^{r+1}=\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}+\rho\boldsymbol{I}\right)^{-1}\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{Y}-\boldsymbol{G}^r+\rho\boldsymbol{Z}^r-\boldsymbol{W}^r\right), \tag{20}$$

requiring to calculate the inverse of a positive definite matrix $\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}+\rho\boldsymbol{I}\right)\in\mathbb{R}^{2J\times 2J}$. However, when the dimensionality of the matrix is large, the computational cost can be substantial. To address this issue, we exploit the sparse matrix Cholesky decomposition technique to approximate $\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}+\rho\boldsymbol{I}\right)^{-1}$, and we define $\mathcal{F}^r=\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{Y}-\boldsymbol{G}^r+\rho\boldsymbol{Z}^r-\boldsymbol{W}^r$. Then, Eq. (20) can be converted to a more computationally efficient form, which can be written as follows:

$$\widehat{\boldsymbol{H}}^{r+1}=\frac{\mathcal{F}^r}{\rho}-\frac{\boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{U}^{-1}\left(\boldsymbol{V}^{-1}\left(\boldsymbol{\Phi}\mathcal{F}^r\right)\right)\right)}{\rho^2}, \tag{21}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are Chorisky decompositions of $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}+\rho\boldsymbol{I}$, i.e., $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}+\rho\boldsymbol{I}=\boldsymbol{U}\boldsymbol{V}$, $\boldsymbol{U}$ is the lower triangular matrix, and $\boldsymbol{V}=\boldsymbol{U}^{\mathrm{T}}$.

2. Update the variable $\boldsymbol{Z}$. Through fixing $\widehat{\boldsymbol{H}}$ and $\boldsymbol{W}$ and applying some operations, the minimization of $\boldsymbol{Z}$ associated with $\mathcal{L}_\rho\left(\widehat{\boldsymbol{H}},\boldsymbol{Z},\boldsymbol{W}\right)$ is represented as

$$\begin{aligned}&\boldsymbol{Z}^{r+1}\\&=\arg\min_{\boldsymbol{Z}}\mathcal{L}_\rho\left(\widehat{\boldsymbol{H}}^{r+1},\boldsymbol{Z},\boldsymbol{W}^r\right)\\&=\arg\min_{\boldsymbol{Z}}\Big(\lambda\left\|\boldsymbol{Z}\right\|_{2,1}+(\boldsymbol{W}^r)^{\mathrm{T}}(\widehat{\boldsymbol{H}}^{r+1}-\boldsymbol{Z})\\&\quad+\frac{\rho}{2}\left\|\widehat{\boldsymbol{H}}^{r+1}-\boldsymbol{Z}\right\|_{\mathrm{F}}^2\Big)\\&=\arg\min_{\boldsymbol{Z}}\Big(\lambda\left\|\boldsymbol{Z}\right\|_{2,1}+\frac{\rho}{2}\left\|\boldsymbol{Z}-\left(\widehat{\boldsymbol{H}}^{r+1}+\boldsymbol{W}^r\right)\right\|_{\mathrm{F}}^2\Big).\end{aligned}\tag{22}$$

There exists a closed-form resolution regarding the matrix $\|\boldsymbol{Z}\|_{2,1}$ norm, and the optimal solution of $\boldsymbol{Z}$ is denoted by

$$\boldsymbol{Z}^{r+1}=\text{Row\_Shrink}\left(\widehat{\boldsymbol{H}}^{r+1}+\boldsymbol{W}^r/\rho,\lambda/\rho\right).\tag{23}$$

$\boldsymbol{Z}^*=\text{Row\_Shrink}\left(\boldsymbol{C},\lambda/\rho\right)$ is the row shrinkage threshold function, $\boldsymbol{Z}^*$ is the optimal solution of $\boldsymbol{Z}$, and $\boldsymbol{C}^r=\widehat{\boldsymbol{H}}^{r+1}+\boldsymbol{W}^r/\rho$. The detailed proof procedure can be found in Lu et al. (2011), and the specific expression is

$$\begin{cases}\left(\boldsymbol{z}_j\right)^{r+1}=\max\left\{\frac{\left\|\boldsymbol{c}_j^r\right\|_2-\lambda/\rho}{\left\|\boldsymbol{c}_j^r\right\|_2}\boldsymbol{c}_j^r,\boldsymbol{0}\right\},\\\boldsymbol{c}_j^r=\left(\widehat{\boldsymbol{h}}_j\right)^{r+1}+\left(\boldsymbol{w}_j\right)^r/\rho,\end{cases}\tag{24}$$

where $\boldsymbol{z}_j$, $\boldsymbol{c}_j$, $\widehat{\boldsymbol{h}}_j$, and $\boldsymbol{w}_j$ are the j^{th} row of $\boldsymbol{Z}$, $\boldsymbol{C}$, $\widehat{\boldsymbol{H}}$, and $\boldsymbol{W}$, respectively.

3. Update the Lagrangian multiplier $\boldsymbol{W}$:

$$\boldsymbol{W}^{r+1}=\boldsymbol{W}^r+\rho\left(\widehat{\boldsymbol{H}}^{r+1}-\boldsymbol{Z}^{r+1}\right).\tag{25}$$

Remark 1 In accordance with the definitions given in Boyd et al. (2011), we set the stop iteration condition for the ADMM solver in Algorithm 1 as follows:

$$\|\widehat{\boldsymbol{H}}^r-\boldsymbol{Z}^r\|_{\mathrm{F}}\le\sqrt{J}\xi^{\text{abs}}+\varsigma^{\text{rel}}\max\{\|\widehat{\boldsymbol{H}}^r\|_{\mathrm{F}},\|\boldsymbol{Z}^r\|_{\mathrm{F}}\},\tag{26}$$

$$\|\rho(\boldsymbol{Z}^r-\boldsymbol{Z}^{r-1})\|_{\mathrm{F}}\le\sqrt{J}\xi^{\text{abs}}+\varsigma^{\text{rel}}\|\boldsymbol{W}^r\|_{\mathrm{F}},\tag{27}$$

where $\widehat{\boldsymbol{H}}^r-\boldsymbol{Z}^r$ and $\boldsymbol{Z}^r-\boldsymbol{Z}^{r-1}$ are the original and dual residuals of the r^{th} iteration respectively, and $\xi^{\text{abs}}>0$ and $\varsigma^{\text{rel}}>0$ are absolute and relative tolerances, respectively. Our solution to problem (12) through the DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm can be summarized as Algorithm 1, where $s_{\max}$ and $r_{\max}$ are the maximum numbers of external and internal iterations, respectively.

Algorithm 1 DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization for solving problem (12)

Initialization: sensing matrix $\boldsymbol{\Phi}$, received signal $\boldsymbol{Y}$, penalty parameter ρ, and tuning parameters λ and γ
Output: estimated channel matrix $\widehat{\boldsymbol{H}}$

while $s<s_{\max}$ **do**
  give the subgradient of $\boldsymbol{G}^s$ according to Eq. (15)
  while $r<r_{\max}$ **do**
    update channel matrix $\widehat{\boldsymbol{H}}$ according to Eq. (21)
    update auxiliary matrix $\boldsymbol{Z}$ according to Eq. (23)
    update Lagrangian multiplier $\boldsymbol{W}$ according to Eq. (25)
    if the stop iteration condition is satisfied **then**
      break
    end if
    $r=r+1$
  end while
  if $\frac{\left\|\widehat{\boldsymbol{H}}^{s+1}-\widehat{\boldsymbol{H}}^s\right\|_{\mathrm{F}}}{\max\left\{\left\|\widehat{\boldsymbol{H}}^s\right\|_{\mathrm{F}},1\right\}}<\gamma$ **then**
    break
  end if
  $s=s+1$
end while

4 Fast DCA-based JADCE algorithm

We employ DCA to solve the MMV-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization problem, and then use an ADMM solver to handle the convex subproblems decomposed by DCA, resulting in a computationally expensive procedure. To alleviate this issue, we derive the analytical solution of the $\mathcal{L}_{2,1-\mathrm{F}}$-metric proximal operator, enabling the ADMM algorithm to solve the optimization problem directly. The solution to problem (12) by the fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization approach is presented as Algorithm 2.

We present the closed-form solution of the fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ proximal operator.

Lemma 1 Consider the optimization problem as follows:

$$\boldsymbol{X}^*=\arg\min_{\boldsymbol{X}}\left(\lambda\left(\|\boldsymbol{X}\|_{2,1}-\beta\|\boldsymbol{X}\|_{\mathrm{F}}\right)+\frac{1}{2}\|\boldsymbol{X}-\boldsymbol{E}\|_{\mathrm{F}}^2\right),\tag{28}$$

where $\boldsymbol{X}^*\in\mathbb{R}^{m\times n}$ is the optimal solution with respect to $\boldsymbol{X}\in\mathbb{R}^{m\times n}$, $\boldsymbol{E}\in\mathbb{R}^{m\times n}$ is a constant matrix, and $\lambda>0$. The minimum value of $\boldsymbol{X}^*$ is expressed

Algorithm 2 Fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization for solving problem (12)

Initialization: sensing matrix $\boldsymbol{\Phi}$, received signal $\boldsymbol{Y}$, penalty parameter ρ, and tuning parameters λ and γ
Output: estimated channel matrix $\widehat{\boldsymbol{H}}$
1: **while** $g < g_{\max}$ **do**
2: update channel matrix $\widehat{\boldsymbol{H}}$ according to Eq. (35)
3: update auxiliary matrix $\boldsymbol{Z}$ according to Eq. (37)
4: update Lagrangian multiplier $\boldsymbol{W}$ according to Eq. (34)
5: **if** $\frac{\left\|\widehat{\boldsymbol{H}}^{g+1}-\widehat{\boldsymbol{H}}^{g}\right\|_{\mathrm{F}}}{\left\|\widehat{\boldsymbol{H}}^{g}\right\|_{\mathrm{F}}}<\gamma$ **then**
6: break
7: **end if**
8: $g=g+1$
9: **end while**

by

$$\boldsymbol{X}^{*}=\mathrm{Row\_Shrink}_{\mathcal{L}_{2,1-\mathrm{F}}}\left(\boldsymbol{E},\lambda\right), \tag{29}$$

where $\mathrm{Row\_Shrink}_{\mathcal{L}_{2,1-\mathrm{F}}}$ is defined as follows:

$$\left(\boldsymbol{x}^{*}\right)_{m}=\begin{cases}\frac{\|\boldsymbol{e}_m\|_2-(1-\beta)\lambda}{\|\boldsymbol{c}_m\|_2}\boldsymbol{c}_m, & \text{if } \|\boldsymbol{e}_m\|_2>\lambda,\\ \boldsymbol{0}, & \text{otherwise.}\end{cases} \tag{30}$$

Proof We define the matrix $(2,1)$-norm and F-norm as $\|\boldsymbol{X}\|_{2,1}=\sum_{i=1}^{m}\|\boldsymbol{x}_i\|_2$ and $\|\boldsymbol{X}\|_{\mathrm{F}}=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{x}_{i,j}^2}$, respectively (Lu et al., 2011). Therefore, the original optimization problem (28) for the fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ proximal operator is equivalent to successive row optimization problems with respect to $\boldsymbol{X}$ and $\boldsymbol{E}$, which can be expressed as follows:

$$\begin{aligned}&\left(\boldsymbol{x}^{*}\right)_{m}\\&=\arg\min_{\boldsymbol{x}}\left(\lambda\left(\|\boldsymbol{x}_m\|_2-\beta\|\boldsymbol{x}_m\|_2\right)+\frac{1}{2}\|\boldsymbol{x}_m-\boldsymbol{e}_m\|_2^2\right)\\&=\arg\min_{\boldsymbol{x}}\left((1-\beta)\lambda\|\boldsymbol{x}_m\|_2+\frac{1}{2}\|\boldsymbol{x}_m-\boldsymbol{e}_m\|_2^2\right),\end{aligned} \tag{31}$$

where $\boldsymbol{x}_m$ and $\boldsymbol{e}_m$ are the m^{th} row of $\boldsymbol{X}$ and $\boldsymbol{E}$, respectively. We define the function $f\left(\boldsymbol{x}_m\right)=(1-\beta)\lambda\|\boldsymbol{x}_m\|_2+\frac{1}{2}\|\boldsymbol{x}_m-\boldsymbol{e}_m\|_2^2$.

1. If $\|\boldsymbol{e}_m\|_2^2\le(1-\beta)\lambda$, assuming $\boldsymbol{x}_m=\theta\boldsymbol{e}_m$, where θ is a scalar parameter, then $f\left(\theta\boldsymbol{e}_m\right)=(1-\beta)\lambda\|\theta\boldsymbol{e}_m\|_2+\frac{1}{2}\|(\theta-1)\boldsymbol{e}_m\|_2^2\ge\frac{1}{2}\left(\theta^2+1\right)\|\boldsymbol{e}_m\|_2^2$. It means when $\theta=0$, $\boldsymbol{e}_m=\boldsymbol{0}$. In this case, the objective function reaches the minimum value.

2. If $\|\boldsymbol{e}_m\|_2^2>(1-\beta)\lambda$, assuming $\boldsymbol{x}_m=\theta\boldsymbol{e}_m$, then $f\left(\theta\boldsymbol{e}_m\right)=\left((1-\beta)\lambda\theta+\frac{1}{2}(\theta-1)^2\|\boldsymbol{e}_m\|_2\right)\|\boldsymbol{e}_m\|_2$. If the minimum value of the objective function $f\left(\theta\boldsymbol{e}_m\right)$ is reached at θ_0, then θ_0 must be a smoothing point. Therefore, let the derivative function of $f\left(\theta\boldsymbol{e}_m\right)$ with respect to θ be 0. Then $\theta_0=\frac{\|\boldsymbol{e}_m\|_2-(1-\beta)\lambda}{\|\boldsymbol{e}_m\|_2}$. The proof is complete.

Consequently, by employing the proximal operation of $\mathcal{L}_{2,1-\mathrm{F}}$, it is possible to solve the objective optimization problem (12) with the ADMM algorithm in a straightforward manner. Similarly, with the introduction of auxiliary variable $\boldsymbol{Z}$, the optimization problem is transformed into the following expression:

$$\begin{aligned}\widehat{\boldsymbol{H}}^{g+1}=&\arg\min_{\boldsymbol{H}}\left(\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\|_2^2+\lambda(\|\boldsymbol{Z}\|_{2,1}-\beta\|\boldsymbol{Z}\|_{\mathrm{F}})\right)\\&\text{s.t.}\quad\boldsymbol{H}-\boldsymbol{Z}=\boldsymbol{0},\end{aligned} \tag{32}$$

where g is the iteration index. Then the augmented Lagrangian function of Eq. (32) is expressed as follows:

$$\begin{aligned}\mathcal{T}_{\rho}(\widehat{\boldsymbol{H}},\boldsymbol{Z},\boldsymbol{W})=&\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\|_2^2+\lambda\left(\|\boldsymbol{Z}\|_{2,1}-\beta\|\boldsymbol{Z}\|_{\mathrm{F}}\right)\\&+\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{H}-\boldsymbol{Z})+\frac{\rho}{2}\|\boldsymbol{H}-\boldsymbol{Z}\|_{\mathrm{F}}^2,\end{aligned} \tag{33}$$

and the iterative process is represented as follows:

$$\begin{cases}\widehat{\boldsymbol{H}}^{g+1}=\underset{\boldsymbol{H}}{\operatorname{argmin}}\left(\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{\Phi}\boldsymbol{H}\|_2^2\right.\\\qquad\left.+(\boldsymbol{W}^{g})^{\mathrm{T}}(\boldsymbol{H}-\boldsymbol{Z}^{g})+\frac{\rho}{2}\|\boldsymbol{H}-\boldsymbol{Z}^{g}\|_{\mathrm{F}}^2\right),\\\boldsymbol{Z}^{g+1}=\underset{\boldsymbol{Z}}{\operatorname{argmin}}\left(\lambda(\|\boldsymbol{Z}^{g}\|_{2,1}-\beta\|\boldsymbol{Z}^{g}\|_{\mathrm{F}})\right.\\\qquad\left.+\frac{\rho}{2}\|\boldsymbol{Z}-(\widehat{\boldsymbol{H}}^{g+1}+\boldsymbol{W}^{g})\|_{\mathrm{F}}^2\right),\\\boldsymbol{W}^{g+1}=\boldsymbol{W}^{g}+\rho\left(\widehat{\boldsymbol{H}}^{g+1}-\boldsymbol{Z}^{g+1}\right).\end{cases} \tag{34}$$

The derivative function of the objective function with respect to $\boldsymbol{H}$ is 0, and the optimal solution for $\widehat{\boldsymbol{H}}$ is expressed as

$$\widehat{\boldsymbol{H}}^{g+1}=\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}+\rho\boldsymbol{I}\right)^{-1}\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{Y}+\rho\boldsymbol{Z}^{g}-\boldsymbol{W}^{g}\right), \tag{35}$$

with the same treatment as Eq. (21). Let $\mathcal{G}^{g}=\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{Y}+\rho\boldsymbol{Z}^{g}-\boldsymbol{W}^{g}$. It is equivalent to converting Eq. (35) to the following equation:

$$\widehat{\boldsymbol{H}}^{g+1}=\frac{\mathcal{G}^{g}}{\rho}-\frac{\boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{U}^{-1}\left(\boldsymbol{V}^{-1}\left(\boldsymbol{\Phi}\mathcal{G}^{g}\right)\right)\right)}{\rho^2}. \tag{36}$$

Based on the closed-form solution of the $\mathcal{L}_{2,1-\mathrm{F}}$ proximal operator, the optimal solution for $\boldsymbol{Z}$ can be expressed as follows:

$$\boldsymbol{Z}^{g+1}=\mathrm{Row\_Shrink}_{\mathcal{L}_{2,1-\mathrm{F}}}\left(\widehat{\boldsymbol{H}}^{g+1}+\boldsymbol{W}^{g}/\rho,\lambda/\rho\right). \tag{37}$$

Remark 2 If the fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm fails to produce an acceptable solution, we apply the adaptive update β strategy (Ge and Li, 2022) for ill-conditioned sensing matrices, and initialize it with $\beta^{(0)}=0.1$. The updating strategy for each β iteration is defined as follows:

$$\beta^{(g+1)}=\begin{cases}\beta^{(g)}, & \text{if } \mod(g,5)\neq 0,\\ \min\left\{1.5\beta^{(g)},1\right\}, & \text{if } \mod(g,5)=0.\end{cases} \tag{38}$$

Remark 3 According to the structured row sparsity property of the estimated channel matrix $\widehat{\boldsymbol{H}}$, we could detect active users according to the segmentation threshold (Shao et al., 2020). Specifically, we identify the estimated active indicator for the k^{th} user by providing the following definition:

$$\alpha_k=\begin{cases}1, & \text{if } \left\|\widehat{\boldsymbol{H}}\left(kN_{\mathrm{BS}},:\right)\right\|_{\mathrm{F}}^2>vN_{\mathrm{BS}},\\ 0, & \text{if } \left\|\widehat{\boldsymbol{H}}\left(kN_{\mathrm{BS}},:\right)\right\|_{\mathrm{F}}^2\leq vN_{\mathrm{BS}},\end{cases} \tag{39}$$

where $v=v_1\max\widehat{\boldsymbol{H}}(j,p)$, $j=1,2,\cdots,2J$, $p=1,2,\cdots,P$, and $\max\widehat{\boldsymbol{H}}(j,p)$ represents the operation of extracting the maximum value of the elements from matrix $\widehat{\boldsymbol{H}}$. The value of v is set to 0.1, the ratio of the minimum amplitude to the maximum amplitude of the channel coefficients.

5 Computational complexity analysis

In the large-scale connectivity scenario of massive multipanel MIMO systems, it is crucial to analyze the effect of algorithm complexity on hardware cost and power consumption. The DCA and fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ algorithms in Table 1 are presented with the number of real multiplications required for each iteration of JADCE, and the other compared algorithms are presented with the number of complex multiplications. The distributed sparse adaptive matching pursuit (DSAMP) algorithm (Gao et al., 2015) involves the computation of the matrix inverse for least-square estimation, and its computational complexity is directly proportional to the cubic of the number of active users K_α. In contrast, the MMV-ADMM algorithm (Lu et al., 2011) requires the matrix inverse in the channel matrix computation, and its computational complexity is proportional to the cubic of the product of the number of users and the number of antennas. On the other hand, the $\mathcal{L}_{2,1-\mathrm{F}}$ algorithms based on DCA and fast DCA do not perform the matrix inverse operation, having significantly lower algorithmic complexity than the DSAMP and MMV-ADMM algorithms, and slightly higher complexity than the section-wise AMP algorithm (Tang et al., 2020). In conclusion, our proposed algorithms are excellent in active user detection and channel estimation and are competitive in computational complexity, and thus are suitable for practical applications in massive multipanel MIMO systems.

6 Simulation results

In this section, we present extensive simulations to evaluate the DCA-based JADCE algorithms for mmWave massive MIMO-OFDM systems in the mMTC scenarios. The parameter settings of the system model used for analysis and simulation are given in Table 2. To facilitate comparison with the two proposed algorithms, we set $\beta=1$ in the DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm, and $0<\beta<1$ in the fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization algorithm, which is the adaptive weight parameter strategy, resulting in a little performance gap between Algorithms 1 and 2, and the other parameters as ρ=1e-3 and $\lambda=$ 1e-2. Finally, each simulation is repeated 100 times. According to the parameters of our simulations, the transmitted delay of an OFDM symbol is equivalent to 0.288 μs.

Then, we use the active detection error probability (ADEP) to evaluate the effectiveness of active user detection algorithms. ADEP includes the probability of missed detection and false detection resulting from the optimization algorithm during the detection of active users (Shao et al., 2020). The mean-squared error (MSE) is exploited to measure the accuracy of the optimization algorithm for recovering the channel matrix. The ADEP and MSE are defined as follows:

$$\mathrm{ADEP}=\frac{1}{K}\sum_{k=1}^{K}|\hat{\alpha}_k-\alpha_k|, \tag{40}$$

$$\mathrm{MSE}=\frac{1}{KN_{\mathrm{BS}}P}\left\|\widehat{\boldsymbol{H}}-\boldsymbol{H}\right\|_{\mathrm{F}}^2. \tag{41}$$

We compare the two proposed MMV-based minimization algorithms with three benchmark algorithms. Section-wise AMP (Tang et al., 2020) is a state-of-the-art AMP algorithm with nonseparable denoisers, where the channel power and the noise

Table 1 Computational complexity of JADCE

Algorithm	Number of complex/real multiplications in each iteration
DSAMP	$(2Q+3)JP+2Q(P+1)t^2+P+2t^3$
MMV-ADMM	$JPQ+J^3$
Section-wise AMP	$2JPQ$
DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization	$r_{\max}\left(8JQ^2+4JPQ\right)$
Fast DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ minimization	$8JQ^2+4JPQ$

t denotes the stage index, and $r_{\max}$ is the maximum number of iterations for internal iterations

Table 2 Parameters used for analysis and simulation

Parameter	Symbol	Value
Number of subarray panels in the horizontal dimension of the multipanel antenna array	I_{h}	4
Number of subarray panels in the vertical dimension of the multipanel antenna array	I_{v}	4
Number of antenna array panels	N_{p}	16
Number of antennas in the horizontal dimension of each subarray panel	M_{h}	2
Number of antennas in the vertical dimension of each subarray panel	M_{v}	2
Time delay of the l^{th} path for the k^{th} user in the t^{th} OFDM symbol	$\tau_{k,l}^{t}$	$\mathcal{U}[0,32/B_{\mathrm{s}}]$
Number of antennas of each subarray panel	N_{BS}	64
Interval of adjacent panels	Δ	30
Number of users	K	500
Number of active users	K_{α}	50
Number of pilot subcarriers	P	16
Number of channel paths	L	4
Number of subcarriers	N_{c}	256
Active factor	P_{α}	0.1
Carrier frequency	f_{c}	30 GHz
System bandwidth	B_{s}	1 GHz

variance are imposed as prior information, and the damping parameter is set to 0.2. The MMV-ADMM algorithm (Lu et al., 2011) is employed to tackle the joint sparse signal recovery problem with the MMV framework, and it adopts the same parameter settings as the proposed algorithms. The DSAMP algorithm (Gao et al., 2015) is designed to jointly estimate the channel problem for multiple subcarriers, where the stopping iteration condition is that the average energy of the channel is lower than the noise power. Notably, the OAMP algorithm (Xiu et al., 2023) can perform well when the sensing matrix leverages a partial unitary matrix but is insensitive to the Gaussian matrix. Therefore, we do not compare the proposed algorithms with the OAMP algorithm.

The MSE performance evaluation of different algorithms versus the number of OFDM symbols varying from 50 to 300 is given in Fig. 3. As the number of OFDM symbols increases, it is evident that our proposed methods show superior performance compared to baseline methods. This is mainly because our proposed algorithms can give sparse solutions compared with other methods in a strongly coherent communication environment. Moreover, our proposed algorithms enhance the sparsity of the channel compared with the MMV-ADMM algorithm. The section-wise AMP algorithm has a slight performance loss owing to fully exploiting the structured sparsity but easily diverges. The DSAMP algorithm results in poor performance because sparsity is not fully exploited, and the strong coherence scenarios cannot be confronted with. Specifically, for $T=150$ and $\boldsymbol{\Phi}\in\mathbb{R}^{2400\times 16\,000}$, we calculate $\mu(\boldsymbol{\Phi})=0.985\,72$, and it means that the measurement matrix has strong coherence in mMTC circumstance. The MSE performance of Algorithm 2 is almost 0.5 orders of magnitude superior to the MSE performance of the MMV-ADMM algorithm.

In Fig. 4, we further present the ADEP performance of various algorithms versus the number of OFDM symbols. Note that MMV-ADMM and DSAMP algorithms exploit the active user detection method described in Remark 3, and that the section-wise AMP algorithm uses the active user detection scheme mentioned in Ke et al. (2020) and Xiu et al.

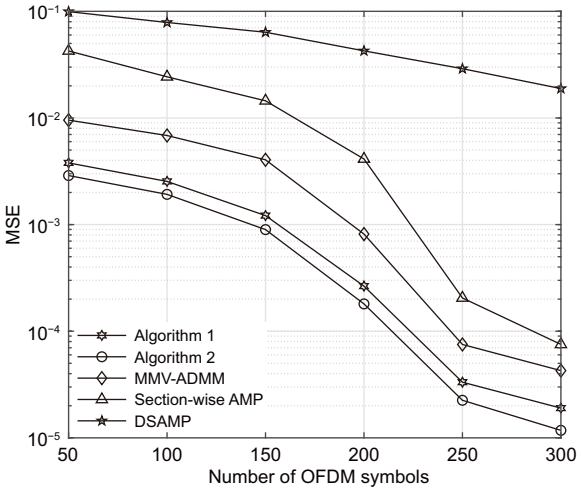

Fig. 3 Channel estimation performance of MSE versus the number of OFDM symbols. The parameter settings are $K = 500$, $K_\alpha = 50$, SNR=30 dB, and $P = 16$

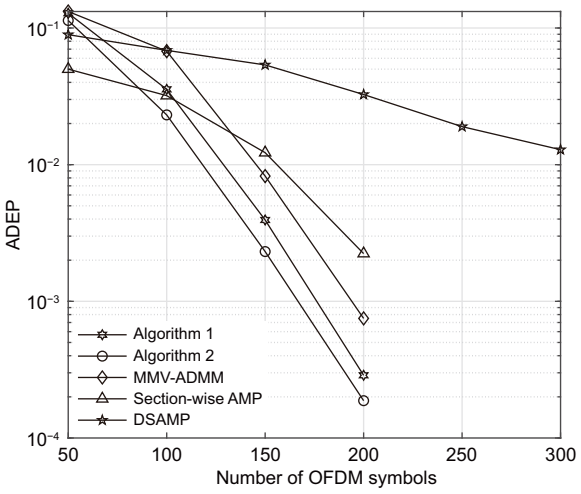

Fig. 4 Active user detection performance of ADEP versus the number of OFDM symbols. The parameter settings are $K = 500$, $K_\alpha = 50$, SNR=30 dB, and $P = 16$

(2023). It is observed that the two proposed algorithms outperform the other compressive sensing algorithms. Specifically, the proposed algorithms, MMV-ADMM, and section-wise AMP can detect the active users with a high probability of success when the number of OFDM symbols is large enough, i.e., $T > 200$, and the ADEP tends to zero rapidly. Furthermore, when $T = 200$, the ADEP performance of the proposed Algorithm 2 is almost 0.5 orders of magnitude superior than that of the MMV-ADMM approach. As a result, it can correctly detect all users with an access delay of 57.6 μs.

Fig. 5 illustrates the MSE performance of different schemes against the transmitting signal-to-noise ratio (SNR), which is expressed as $\text{SNR} \triangleq 10\lg\left(\frac{\gamma}{\sigma^2}\right)$, where γ is the transmitting power. It can be observed that the performances of all the algorithms increase gradually with the increase of transmitting SNR. This is attributed to the fact that when the SNR is gradually increased, the received signal is less affected by the noise; thus, the estimated channel matrix is closer to the actual one. Furthermore, compared with other approaches, the performance gains of the two proposed methods consistently show significant advantages even though the transmitting power is low. In other words, it is well illustrated that Algorithms 1 and 2 have better robustness to noise.

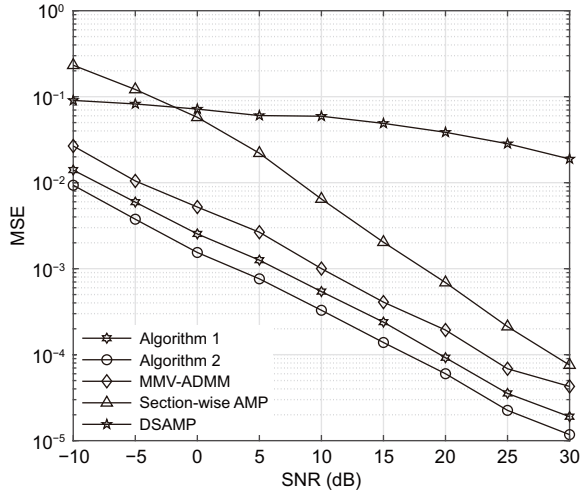

Fig. 5 Channel estimation performance of MSE versus SNR. The parameter settings are $K = 500$, $K_\alpha = 50$, $T = 300$, and $P = 16$

The MSE performance evaluation of the proposed algorithms and state-of-the-art benchmark methods versus the number of pilot subcarriers is given in Fig. 6. It is observed that the performances of all the algorithms are improved steadily as the number of pilot subcarriers P increases, and that our proposed algorithms have superior performance than other approaches. Specifically, when $P = 4$, Algorithm 2 performs similarly to the MMV-ADMM algorithm with the pilot subcarrier $P = 10$. This demonstrates that the proposed algorithms can perform well with a tiny pilot overhead.

In Fig. 7, we investigate the proposed algorithms and existing benchmark methods with different active factor values P_α. It is observed that the performances of all the algorithms are degraded as

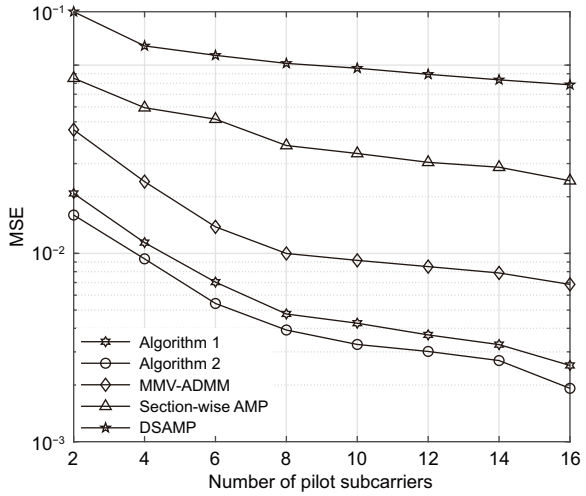

Fig. 6 Channel estimation performance of MSE versus the number of pilot subcarriers. The parameter settings are $K = 500$, $K_\alpha = 50$, $T = 100$, and SNR=30 dB

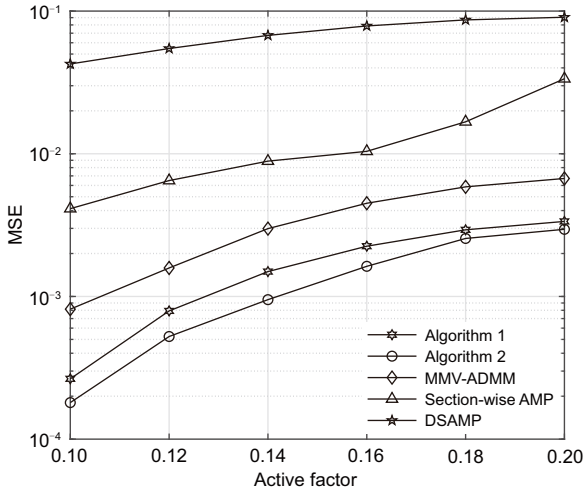

Fig. 7 Channel estimation performance of MSE versus different activity probabilities. The parameter settings are $K = 500$, $T = 200$, SNR=30 dB, and $P = 16$

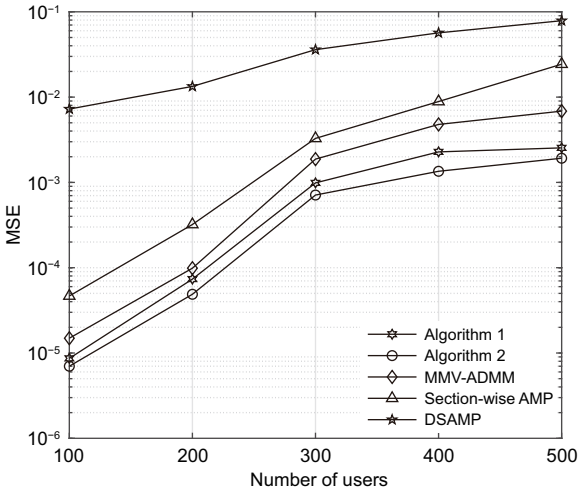

Fig. 8 Channel estimation performance of MSE versus the number of users. The parameter settings are $T = 100$, SNR=30 dB, $P_\alpha = 0.1$, and $P = 16$

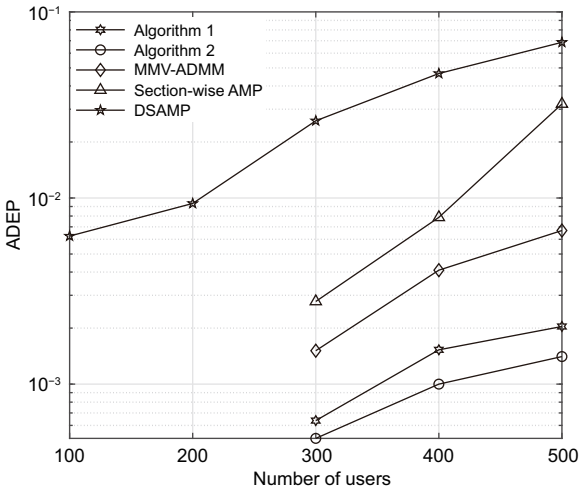

Fig. 9 Active user detection performance of ADEP versus the number of users. The parameter settings are $T = 100$, SNR=30 dB, $P_\alpha = 0.1$, and $P = 16$

P_α increases. This is mainly because the increased number of active users leads to incremental interference between active users. However, the proposed algorithms consistently outperform the three compared algorithms in the MSE performance across the entire range of the active factor. This indicates that our proposed algorithms can satisfy the requirements of various application scenarios of the IoT.

Figs. 8 and 9 compare the MSE and ADEP performances of various schemes with different numbers of potential users, respectively. It is observed that when the number of users is 100 or 200, the proposed algorithms, MMV-ADMM, and section-wise AMP have perfect MSE performance and achieve accurate active user detection. Also, all the users can achieve correct detection at 28.8 μs. Moreover, the performances of all the algorithms gradually degrade as the total number of users increases since it will result in more interference between users. However, comparing results with other classical algorithms, the proposed algorithms can present substantially better MSE and ADEP performances. Therefore, the proposed schemes can accommodate more users in tremendous access systems, making grant-free 6G cellular IoT access more appealing.

7 Conclusions

This article has investigated the JADCE problem in strongly coherent mMTC scenarios based on a grant-free scenario. Due to the strongly coherent communication systems, the performance of existing compressed sensing algorithms degrades as the number of pilot subcarriers declines. To address this issue, we have formulated the JADCE problem as a joint sparse signal recovery problem and introduced a novel DCA-based $\mathcal{L}_{2,1-\mathrm{F}}$ algorithm with the MMV framework to solve the JADCE problem. To further reduce the computational complexity of the DCA-based JADCE algorithm, a fast DCA-based JADCE algorithm has been provided, from which the analytical solution based on the $\mathcal{L}_{2,1-\mathrm{F}}$-metric proximal operator is derived, and hence the ADMM solver can directly address the optimization problem. Simulation results have demonstrated that the two proposed DC optimization based JADCE algorithms can achieve significant performance gains in MSE and ADEP over conventional sparse recovery algorithms.

Contributors

Kaihui LIU designed the research. Lijun ZHU and Kaihui LIU performed the simulations and analyzed the data. All the authors drafted, revised, and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Alkhateeb A, El Ayach O, Leus G, et al., 2014. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE J Sel Top Signal Process*, 8(5):831-846. https://doi.org/10.1109/JSTSP.2014.2334278

Bian XY, Mao YY, Zhang J, 2023. Joint activity detection, channel estimation, and data decoding for grant-free massive random access. *IEEE Int Things J*, 10(16):14042-14057. https://doi.org/10.1109/JIOT.2023.3243947

Bian XY, Mao YY, Zhang J, 2024. Joint activity-delay detection and channel estimation for asynchronous massive random access: a free probability theory approach. https://arxiv.org/abs/2402.17996

Boyd S, Parikh N, Chu E, et al., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found Trends Mach Learn*, 3(1):1-122. https://doi.org/10.1561/2200000016

Chukhno N, Chukhno O, Moltchanov D, et al., 2024. Models, methods, and solutions for multicasting in 5G/6G mmWave and sub-THz systems. *IEEE Commun Surv Tutor*, 26(1):119-159. https://doi.org/10.1109/COMST.2023.3319354

Cui Y, Li SC, Zhang WQ, 2021. Jointly sparse signal recovery and support recovery via deep learning with applications in MIMO-based grant-free random access. *IEEE J Sel Areas Commun*, 39(3):788-803. https://doi.org/10.1109/JSAC.2020.3018802

Djelouat H, Leinonen M, Juntti M, 2022. Spatial correlation aware compressed sensing for user activity detection and channel estimation in massive MTC. *IEEE Trans Wirel Commun*, 21(8):6402-6416. https://doi.org/10.1109/TWC.2022.3149111

Gan X, Zhong CJ, Huang CW, et al., 2021. RIS-assisted multi-user MISO communications exploiting statistical CSI. *IEEE Trans Commun*, 69(10):6781-6792. https://doi.org/10.1109/TCOMM.2021.3100860

Gao Z, Dai LL, Wang ZC, et al., 2015. Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO. *IEEE Trans Signal Process*, 63(23):6169-6183. https://doi.org/10.1109/TSP.2015.2463260

Gao Z, Ke ML, Mei YK, et al., 2024. Compressive sensing-based grant-free massive access for 6G massive communication. *IEEE Int Things J*, 11(5):7411-7435. https://doi.org/10.1109/JIOT.2023.3334878

Ge HM, Li P, 2022. The Dantzig selector: recovery of signal via $\ell_1-\alpha\ell_2$ minimization. *Inv Probl*, 38(1):015006. https://doi.org/10.1088/1361-6420/ac39f8

Guo MQ, Gursoy MC, 2023. Joint activity detection and channel estimation for intelligent-reflecting-surface-assisted wireless IoT networks. *IEEE Int Things J*, 10(12):10207-10221. https://doi.org/10.1109/JIOT.2023.3238972

Guo YR, Liu ZJ, Sun YJ, 2024. Low-complexity joint activity detection and channel estimation with partially orthogonal pilot for asynchronous massive access. *IEEE Int Things J*, 11(1):1773-1783. https://doi.org/10.1109/JIOT.2023.3290976

Ke ML, Gao Z, Wu YP, et al., 2020. Compressive sensing-based adaptive active user detection and channel estimation: massive access meets massive MIMO. *IEEE Trans Signal Process*, 68:764-779. https://doi.org/10.1109/TSP.2020.2967175

Li S, Xiao LX, Jiang T, 2021. An efficient matching pursuit based compressive sensing detector for uplink grant-free NOMA. *IEEE Trans Veh Technol*, 70(2):2012-2017. https://doi.org/10.1109/TVT.2021.3056462

Li Y, Chen SY, Meng WX, et al., 2024. Correlation aided joint activity detection and channel estimation for multi-device collaborative massive access. *IEEE Int*

Things J, 11(10):18394-18409. https://doi.org/10.1109/JIOT.2024.3363704

Liu KH, Li XJ, Fang J, et al., 2019. Bayesian mmWave channel estimation via exploiting joint sparse and low-rank structures. *IEEE Access*, 7:48961-48970. https://doi.org/10.1109/ACCESS.2019.2910088

Liu KH, Wan LT, Sun L, 2022. Fast quadratic sensing via nonconvex optimization. Proc 30$^{\text{th}}$ European Signal Processing Conf, p.2211-2215. https://doi.org/10.23919/eusipco55093.2022.9909632

Liu KH, Li XN, Zhao HY, et al., 2023. Joint active user detection and channel estimation for massive grant-free access via difference of convex programming. Proc IEEE Global Communications Conf, p.2335-2340. https://doi.org/10.1109/globecom54140.2023.10437881

Liu L, Yu W, 2018. Massive connectivity with massive MIMO–part I: device activity detection and channel estimation. *IEEE Trans Signal Process*, 66(11):2933-2946. https://doi.org/10.1109/TSP.2018.2818082

Lou YF, Yan M, 2018. Fast L_1-L_2 minimization via a proximal operator. *J Sci Comput*, 74(2):767-785. https://doi.org/10.1007/s10915-017-0463-2

Lou YF, Yin PH, He Q, et al., 2015. Computing sparse representation in a highly coherent dictionary based on difference of L_1 and L_2. *J Sci Comput*, 64(1):178-196. https://doi.org/10.1007/s10915-014-9930-1

Lu HT, Long XZ, Lv JY, 2011. A fast algorithm for recovery of jointly sparse vectors based on the alternating direction methods. Proc 14$^{\text{th}}$ Int Conf on Artificial Intelligence and Statistics, p.461-469.

Ma Z, Wu W, Gao FF, et al., 2024. Model-driven deep learning for non-coherent massive machine-type communications. *IEEE Trans Wirel Commun*, 23(3):2197-2211. https://doi.org/10.1109/TWC.2023.3296218

Marata L, López OLA, Hauptmann A, et al., 2023. Joint activity detection and channel estimation for clustered massive machine type communications. *IEEE Trans Wirel Commun*, 23(6):5473-5487. https://doi.org/10.1109/TWC.2023.3326468

Mei YK, Gao Z, Mi D, et al., 2023. Massive access in extra large-scale MIMO with mixed-ADC over near-field channels. *IEEE Trans Veh Technol*, 72(9):12373-12378. https://doi.org/10.1109/TVT.2023.3266230

Petersen KB, Pedersen MS, 2012. The Matrix Cookbook. http://www2.imm.dtu.dk/pubdb/edoc/imm3274.pdf

Qiao L, Zhang J, Gao Z, et al., 2022. Joint activity and blind information detection for UAV-assisted massive IoT access. *IEEE J Sel Areas Commun*, 40(5):1489-1508. https://doi.org/10.1109/JSAC.2022.3143255

Rajoriya A, Budhiraja R, 2023. Joint AMP-SBL algorithms for device activity detection and channel estimation in massive MIMO mMTC systems. *IEEE Trans Commun*, 71(4):2136-2152. https://doi.org/10.1109/TCOMM.2023.3244225

Shao XD, Chen XM, Jia RD, 2020. A dimension reduction-based joint activity detection and channel estimation algorithm for massive access. *IEEE Trans Signal Process*, 68:420-435. https://doi.org/10.1109/TSP.2019.2961299

Tang ZH, Wang J, Wang JT, et al., 2020. Device activity detection and non-coherent information transmission for massive machine-type communications. *IEEE Access*, 8:41452-41465. https://doi.org/10.1109/ACCESS.2020.2976824

Tong X, Zhang ZY, Wang J, et al., 2021. Joint multi-user communication and sensing exploiting both signal and environment sparsity. *IEEE J Sel Top Signal Process*, 15(6):1409-1422. https://doi.org/10.1109/JSTSP.2021.3111432

Wan LT, Liu KH, Zhang W, 2022. Deep learning-aided off-grid channel estimation for millimeter wave cellular systems. *IEEE Trans Wirel Commun*, 21(5):3333-3348. https://doi.org/10.1109/TWC.2021.3120926

Wang BC, Dai LL, Mir T, et al., 2016. Joint user activity and data detection based on structured compressive sensing for NOMA. *IEEE Commun Lett*, 20(7):1473-1476. https://doi.org/10.1109/LCOMM.2016.2560180

Wang W, Zhang W, Li YJ, et al., 2018. Channel estimation and hybrid precoding for multi-panel millimeter wave MIMO. Proc IEEE Int Conf on Communications, p.1-6. https://doi.org/10.1109/ICC.2018.8422137

Wei L, Huang CW, Guo QH, et al., 2022. Joint channel estimation and signal recovery for RIS-empowered multiuser communications. *IEEE Trans Commun*, 70(7):4640-4655. https://doi.org/10.1109/TCOMM.2022.3179771

Xiu HL, Gao Z, Liao AW, et al., 2023. Joint activity detection and channel estimation for massive IoT access based on millimeter-wave/terahertz multi-panel massive MIMO. *IEEE Trans Veh Technol*, 72(1):1349-1354. https://doi.org/10.1109/TVT.2022.3206492

Yin PH, Lou YF, He Q, et al., 2015. Minimization of ℓ_1-2 for compressed sensing. *SIAM J Sci Comput*, 37(1):A536-A563. https://doi.org/10.1137/140952363

Ying KK, Gao Z, Chen S, et al., 2023. Quasi-synchronous random access for massive MIMO-based LEO satellite constellations. *IEEE J Sel Areas Commun*, 41(6):1702-1722. https://doi.org/10.1109/JSAC.2023.3273699

Yu KW, Shen M, Wang R, et al., 2020. Joint nuclear norm and ℓ_{1-2}-regularization sparse channel estimation for mmWave massive MIMO systems. *IEEE Access*, 8:155409-155416. https://doi.org/10.1109/ACCESS.2020.3019269

Zhang XX, Labeau F, Hao L, et al., 2021. Joint active user detection and channel estimation via Bayesian learning approaches in MTC communications. *IEEE Trans Veh Technol*, 70(6):6222-6226. https://doi.org/10.1109/TVT.2021.3077569

Zhang XX, Fan PZ, Hao L, et al., 2023. Generalized approximate message passing based Bayesian learning detectors for uplink grant-free NOMA. *IEEE Trans Veh Technol*, 72(11):15057-15061. https://doi.org/10.1109/TVT.2023.3280919

Zhang YY, Guo QH, Wang ZY, et al., 2018. Block sparse Bayesian learning based joint user activity detection and channel estimation for grant-free NOMA systems. *IEEE Trans Veh Technol*, 67(10):9631-9640. https://doi.org/10.1109/TVT.2018.2859806

Zhang ZJ, Li Y, Huang CW, et al., 2019. DNN-aided block sparse Bayesian learning for user activity detection and channel estimation in grant-free non-orthogonal random access. *IEEE Trans Veh Technol*, 68(12):12000-12012. https://doi.org/10.1109/TVT.2019.2947214

Zhang ZJ, Guo QH, Li Y, et al., 2023. Variational Bayesian inference clustering-based joint user activity and data detection for grant-free random access in mMTC. *IEEE Int Things J*, 10(11):9906-9916. https://doi.org/10.1109/JIOT.2023.3234691

Zheng ST, Wu S, Jia HG, et al., 2024. Hybrid driven learning for joint activity detection and channel estimation in IRS-assisted massive connectivity. *IEEE Trans Wirel Commun*, 23(9):10834-10849. https://doi.org/10.1109/TWC.2024.3376381

Zhu LJ, Liu KH, Wan LT, et al., 2023. Active user detection and channel estimation via fast ADMM. Proc IEEE Wireless Communications and Networking Conf, p.1-6. https://doi.org/10.1109/WCNC55385.2023.10118629

Appendix: Proof of the convergence of DCA

We concisely demonstrate that the iterative DCA in Eq. (13) produces a monotonically decreasing objective function. According to the definition of the subgradient, we have

$$\mathcal{W}(\boldsymbol{H}) \geq \mathcal{W}\left(\boldsymbol{H}^s\right) + \left\langle \boldsymbol{Z}^s, \boldsymbol{H} - \boldsymbol{H}^s \right\rangle, \quad \forall \boldsymbol{H} \in \mathbb{R}^s.$$

Specifically, $\mathcal{W}\left(\boldsymbol{H}^{s+1}\right) \geq \mathcal{W}\left(\boldsymbol{H}^s\right) + \left\langle \boldsymbol{Z}^s, \boldsymbol{H}^{s+1} - \boldsymbol{H}^s \right\rangle$, and therefore we have

$$\begin{aligned}
&\mathcal{Q}\left(\boldsymbol{H}^s\right) \\
=&\mathcal{M}\left(\boldsymbol{H}^s\right) - \mathcal{W}\left(\boldsymbol{H}^s\right) \\
&\geq \mathcal{M}\left(\boldsymbol{H}^{s+1}\right) - \left(\mathcal{W}\left(\boldsymbol{H}^s\right) + \left\langle \boldsymbol{Z}^s, \boldsymbol{H}^{s+1} - \boldsymbol{H}^s \right\rangle\right) \\
&\geq \mathcal{M}\left(\boldsymbol{H}^{s+1}\right) - \mathcal{W}\left(\boldsymbol{H}^{s+1}\right) \\
&= \mathcal{Q}\left(\boldsymbol{H}^{s+1}\right).
\end{aligned}$$

The proof is complete.