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

Compressed sensing-based structured joint channel estimation in a multi-user massive MIMO system*

Ruo-yu ZHANG, Hong-lin ZHAO†‡, Shao-bo JIA

(*Communication Research Center, Harbin Institute of Technology, Harbin 150080, China*)

†E-mail: hlzhao@hit.edu.cn

Received Oct. 14, 2016; Revision accepted May 22, 2017; Crosschecked Dec. 20, 2017

Abstract: Acquisition of accurate channel state information (CSI) at transmitters results in a huge pilot overhead in massive multiple input multiple output (MIMO) systems due to the large number of antennas in the base station (BS). To reduce the overwhelming pilot overhead in such systems, a structured joint channel estimation scheme employing compressed sensing (CS) theory is proposed. Specifically, the channel sparsity in the angular domain due to the practical scattering environment is analyzed, where common sparsity and individual sparsity structures among geographically neighboring users exist in multi-user massive MIMO systems. Then, by equipping each user with multiple antennas, the pilot overhead can be alleviated in the framework of CS and the channel estimation quality can be improved. Moreover, a structured joint matching pursuit (SJMP) algorithm at the BS is proposed to jointly estimate the channel of users with reduced pilot overhead. Furthermore, the probability upper bound of common support recovery and the upper bound of channel estimation quality using the proposed SJMP algorithm are derived. Simulation results demonstrate that the proposed SJMP algorithm can achieve a higher system performance than those of existing algorithms in terms of pilot overhead and achievable rate.

Key words: Compressed sensing; Multi-user massive multiple input multiple output (MIMO); Frequency-division duplexing; Structured joint channel estimation; Pilot overhead reduction

https://doi.org/10.1631/FITEE.1601635 **CLC number:** TN911.5

1 Introduction

Massive multiple input multiple output (MIMO), in which base stations (BSs) are equipped with a large number of antennas, has been recognized as a key technique for the next generation wireless systems due to the great gains in capacity and spectrum efficiency it can provide (Lu *et al*., 2014). An essential aim of a massive MIMO system is to acquire the channel state information (CSI) at the BS, so that not only the dramatic array gains due to large spatial degrees of freedom (DoFs) can be fully exploited, but also precoding and beamforming technology can be applied (Bogale *et al*., 2012). One challenge of CSI acquisition is that downlink pilots sent from the BS antennas consume many radio resources due to the large number of antennas (Björnson *et al*., 2015). In time-division duplexing (TDD) massive MIMO systems, channel reciprocity (Hoydis *et al*., 2013), which can be used to acquire CSI by using uplink pilots, greatly reduces the overhead in the process of CSI acquisition. Nevertheless, this feature cannot be used in frequency-division duplexing (FDD) systems (Dai *et al*., 2013; Lee *et al*., 2015), which are now widely used in wireless cellular networks.

Conventional least squares (LS) algorithms and minimum mean square error (MMSE) algorithms (Yin *et al*., 2012) work well in current long term evolution (LTE) systems and LTE-Advanced systems (Ketonen *et al*., 2010) because the number of

‡ Corresponding author

* Project supported by the Fundamental Research Funds for the Central Universities (No. HIT.MKSTISP.2016 13) and the National Natural Science Foundation of China (No. 61671176)

antennas is relatively small. In massive MIMO systems, however, the downlink pilots overhead would be prohibitively large due to the huge number of BS antennas (Noh *et al*., 2014). Hence, it is imperative to design a practical channel estimation algorithm to reduce such a large pilot overhead without lowering channel estimation performance. In recent research, Berger *et al*. (2010) investigated channel sparsity, an important feature of channel in massive MIMO systems, which has been found to be helpful in dealing with the above problems. According to many experimental studies, the channel matrices in the angular domain tend to be sparse as the number of antennas increases due to the limited local scatterers around the BS (Gao *et al*., 2011; Barbotin *et al*., 2012; Hoydis *et al*., 2012). Interestingly, compressed sensing (CS) (Donoho, 2006; Chen and Qin, 2015) can recover the original signal using a small number of measurements with an overwhelming probability under the condition that the original signal is sparse. By exploiting CS, fewer pilots need to be implemented to enhance channel estimation performance. CS was first applied to CSI estimation by using the orthogonal matching pursuit (OMP) (Tropp and Gilbert, 2007) algorithm and OMP-based (Tropp *et al*., 2006; Dai and Milenkovic, 2009) algorithm, achieving better estimation performance than the conventional LS method in terms of the sparse channel. Zhang *et al*. (2014) proposed to condense the CSI into low-dimensional vectors in a feedback process based on CS. Channel sparsity in the delay-domain is then fully exploited to improve channel estimation performance. Dai *et al*. (2013) proposed to exploit the temporal correlation and sparsity of channels in the delay-domain to reduce pilot overhead for FDD massive MIMO systems. Qi and Wu (2014) reported that by virtue of the spatial correlation and channel sparsity of the delay-domain, the pilot overhead to estimate channels of antenna-user links can be reduced. However, these algorithms assume that the users know precisely the channel sparsity level in practical scenarios, which is unrealistic in practical systems. Gao *et al*. (2016) proposed an adaptive structured subspace pursuit algorithm for channel estimation by capitalizing on the spatio-temporal common sparsity of delay-domain MIMO channels. Although the main feature of this algorithm is that it needs no information about accurate channel sparsity as the initial condition, the channel recovery time is too long and it is unsuitable for the high data transfer rate in next generation systems (Hu *et al*., 2013). By using long-term channel statistics, Choi *et al*. (2014) proposed an open-loop and closed-loop channel estimation scheme for FDD massive MIMO systems, but it can be difficult for the users to obtain channel statistics perfectly in practical applications using this scheme. Therefore, there is a need to design a practical algorithm for the BS to estimate the channel matrix.

The channel estimation methods mentioned above apply CS only to point-to-point links, i.e., between the BS and one user, and joint sparsity of channel matrices among geographically neighboring users is not considered. Instead of exploiting only point-to-point channel sparsity, to further improve the channel estimation performance and reduce pilot overhead, multi-user channel matrices in massive MIMO systems are observed to share common local scatterers, and can be recovered jointly in many experimental studies (Gao *et al*., 2011). Rao and Lau (2014) proposed a distributed CS recovery framework exploiting common sparsity between users, which is based on joint sparsity of the channel matrix. A set of sparse channel matrices are distributed, measured, and recovered jointly at the BS. However, a joint OMP algorithm needs corresponding parameters in the process of channel estimation, which makes it difficult to apply to practical scenarios due to changing situations. Gao *et al*. (2014) proposed to leverage common sparsity in large-scale MIMO channels to realize reliable channel estimation and recover multiple channels simultaneously with low pilot overhead. This algorithm exploits the channel sparsity in the time delay domain. Extending such an approach to a multi-user scenario can be inappropriate as the common sparsity among users fails to be used. Based on the spatially common sparsity among the subchannels of the orthogonal frequency division multiplexing (OFDM) symbol within the bandwidth for each user and the unchanged support within the consecutive time blocks, Gao *et al*. (2015) proposed a closed-loop CSI acquisition and feedback frame to adaptively obtain CSI according to the channel measurement of previous time blocks. However, this can be infeasible when the temporal correlation disappears and the approach fails to fully exploit the common sparsity of channel matrices among the

geographically neighboring users. The studies described above concentrate on users with a single antenna, but users can be equipped with more than one antenna to achieve higher transmit rates. Whether multiple antennas can improve channel estimation performance has not yet been investigated.

In this study, by leveraging sparse channels in the angular domain, a structured joint channel estimation scheme is proposed for the BS with significantly less pilot overhead in multi-user FDD massive MIMO systems. Due to the limited scatterers in practical environments, channel matrices associated with neighboring users usually have a structured channel sparsity, including common sparsity and individual sparsity. The multiple antennas of each user are investigated in the framework of CS to further reduce the pilot overhead requirement and improve the channel estimation quality. Instead of using the instantaneous sparsity, the statistical sparsity structure is fully exploited, including common and individual support. Based on this feature, a structured joint matching pursuit (SJMP) algorithm at the BS is proposed to estimate the channel matrix with a smaller pilot overhead. Furthermore, the channel common support recovery probability and channel estimation quality using our proposed SJMP algorithm are analyzed thoroughly. The main contributions of this study are summarized as follows:

Due to the practical scatterer environment, channel matrices associated with neighboring users usually exploit common and individual sparsity, which is called 'structured channel sparsity'. An analysis of the channel property in the angular domain is presented according to the number of antennas in the BS and among users. This is different from the channel sparsity considered by Dai *et al*. (2013), Hu *et al*. (2013), and Gao *et al*. (2014), where the channel sparsity exists in the time delay domain due to multipath signal propagation.

Each user is equipped with multiple antennas, whereas most studies consider that the user has only one antenna. Although Rao and Lau (2014) mentioned this scenario, they did not investigate whether multiple antennas can improve the channel estimation quality. This study proves that the channel estimation performance can be improved by increasing the number of antennas for each user, which provides an insight for designing algorithms for CS-based channel estimation.

Based on the above analysis, an SJMP algorithm at the BS is proposed to estimate the channel matrix with a lower pilot overhead, in which the structured channel sparsity and multiple antennas of users are fully exploited. Furthermore, we analyze the recovery probability of common support and channel estimation quality of the proposed SJMP algorithm, so that the estimated channel matrix can be obtained iteratively to approach the true channel matrix.

Notations: Vectors and matrices are written in lower-case and upper-case boldface, respectively; $|\cdot|$ denotes the cardinality of a set, while $|\cdot|_p$ and $|\cdot|_{\mathrm{F}}$ denote the l_p-norm and Frobenius norm, respectively. The matrix transpose and conjugate transpose are denoted by $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$, respectively, while $(\cdot)^{\dagger}$ is the Moore-Penrose inversion. For a given matrix $\boldsymbol{A}$ and vector $\boldsymbol{a}$, $\boldsymbol{A}(l)$ and $\boldsymbol{a}(l)$ denote the lth column vector of $\boldsymbol{A}$ and lth entry of $\boldsymbol{a}$, respectively. For an index set Γ, $\boldsymbol{A}_{\downarrow\Gamma}$ and $\boldsymbol{A}^{\Gamma\rightarrow}$ denote the sub-matrices consisting of columns and rows, of $\boldsymbol{A}$ whose indices belong to index set Γ, respectively. supp($\boldsymbol{a}$) denotes the index set of non-zero entries of vector $\boldsymbol{a}$. $I_{\{\cdot\}}$ and $\Pr(\cdot)$ denote the indicator function and probability, respectively.

2 Sparse channel model in the angular domain

Consider a single cell massive MIMO system, consisting of one BS and K users. The BS is equipped with M antennas and each user is equipped with N antennas (Fig. 1). All the antennas are assumed to be uniform linear arrays. The ith user channel matrix $\boldsymbol{H}_i$ can be expressed as

$$\boldsymbol{H}_i=\sum_p a_p \boldsymbol{e}_{\mathrm{r}}(\Omega_{\mathrm{r}p})\boldsymbol{e}_{\mathrm{t}}^{\mathrm{H}}(\Omega_{\mathrm{t}p}), \tag{1}$$

where a_p denotes the channel gain of the pth path. $\Omega_{\mathrm{r}p}=\cos\theta_{\mathrm{r}p}$, $\Omega_{\mathrm{t}p}=\cos\theta_{\mathrm{t}p}$, where $\theta_{\mathrm{r}p}$ and $\theta_{\mathrm{t}p}$ represent the degree of arrival and departure of the pth path, respectively. $\boldsymbol{e}_{\mathrm{r}}(\Omega)$ and $\boldsymbol{e}_{\mathrm{t}}(\Omega)$ denote the received and transmitted unit vectors, along the direction Ω, respectively. Specifically,

$$\boldsymbol{e}_{\mathrm{r}}(\Omega)=1/\sqrt{N}\times[1,\ \mathrm{e}^{-\mathrm{j}2\pi\Delta_{\mathrm{r}}\Omega},\ \cdots,\ \mathrm{e}^{-\mathrm{j}2\pi(N-1)\Delta_{\mathrm{r}}\Omega}]^{\mathrm{T}}, \tag{2}$$

$$\boldsymbol{e}_{\mathrm{t}}(\Omega)=1/\sqrt{M}\times[1,\ \mathrm{e}^{-\mathrm{j}2\pi\Delta_{\mathrm{t}}\Omega},\ \cdots,\ \mathrm{e}^{-\mathrm{j}2\pi(M-1)\Delta_{\mathrm{t}}\Omega}]^{\mathrm{T}},\quad(3)$$

where Δ_{r} and Δ_{t} denote the normalized distance between adjacent received antennas and adjacent transmitted antennas, respectively. Ω and Ω' between the BS and a user is denoted by

$$\begin{aligned}f_{\mathrm{r}}(\Delta\Omega)&=f_{\mathrm{r}}(\Omega-\Omega')=\boldsymbol{e}_{\mathrm{r}}^{\mathrm{H}}(\Omega)\boldsymbol{e}_{\mathrm{r}}(\Omega')\\&=\frac{1}{N}\mathrm{e}^{\mathrm{j}\pi(N-1)\Delta_{\mathrm{r}}\Delta\Omega}\frac{\sin(\pi N\Delta_{\mathrm{r}}\Delta\Omega)}{\sin(\pi\Delta_{\mathrm{r}}\Delta\Omega)}.\end{aligned}\quad(4)$$

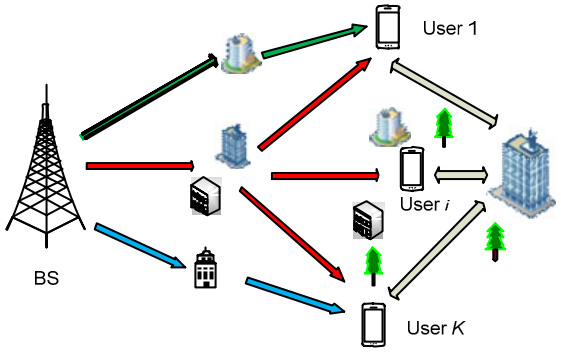

Fig. 1 Multi-user massive MIMO model

It can be seen that $f_{\mathrm{r}}(\Delta\Omega)$ has periodicity, that is,

$$\begin{aligned}&f_{\mathrm{r}}(k/L_{\mathrm{r}})=0,\ f_{\mathrm{r}}(-k/L_{\mathrm{r}})=f_{\mathrm{r}}((N-k)/L_{\mathrm{r}}),\\&k=1,2,\cdots,N-1,\end{aligned}\quad(5)$$

where $L_{\mathrm{r}}=N\Delta_{\mathrm{r}}$. For each user, when the degree of arrival from different directions satisfies Eq. (5), these paths are orthogonal with each other. Thus, an N-dimensional vector can be formed as an orthogonal basis $\boldsymbol{\Theta}_{\mathrm{R}}$:

$$\boldsymbol{\Theta}_{\mathrm{R}}=\{\boldsymbol{e}_{\mathrm{r}}(0),\boldsymbol{e}_{\mathrm{r}}(1/L_{\mathrm{r}}),\cdots,\boldsymbol{e}_{\mathrm{r}}((N-1)/L_{\mathrm{r}})\}.\quad(6)$$

Then the received signal can be denoted in the angular domain using the orthogonal basis in Eq. (6). This indicates that the channel matrix can provide a resolution of $1/L_{\mathrm{r}}$. When the physical paths of the received signal are within the resolution that the antenna array can provide, these different paths will not be distinguished. The paths are regarded as a cluster from the same direction in the view of the receiver. That is to say, the received signal energy from all the physical directions can be distributed into a particular vector $\boldsymbol{e}_{\mathrm{r}}(k/L_{\mathrm{r}})$. Similarly, the transmitted signal can be represented in the orthogonal basis $\boldsymbol{\Theta}_{\mathrm{T}}$:

$$\boldsymbol{\Theta}_{\mathrm{T}}=\{\boldsymbol{e}_{\mathrm{t}}(0),\boldsymbol{e}_{\mathrm{t}}(1/L_{\mathrm{t}}),\cdots,\boldsymbol{e}_{\mathrm{t}}((M-1)/L_{\mathrm{t}})\},\quad(7)$$

where $L_{\mathrm{t}}=M\Delta_{\mathrm{t}}$. Accordingly, the energy of signal transmitted in any physical direction can be denoted by a particular vector $\boldsymbol{e}_{\mathrm{t}}(l/L_{\mathrm{t}})$.

From the analysis above, the channel matrix of the ith user is denoted in the angular domain by

$$\boldsymbol{H}_i^{\mathrm{a}}=\boldsymbol{\Theta}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{H}_i\boldsymbol{\Theta}_{\mathrm{T}},\quad(8)$$

where $\boldsymbol{H}_i^{\mathrm{a}}$ is the channel matrix in the angular domain. Specifically, the element in the kth row and lth column of $\boldsymbol{H}_i^{\mathrm{a}}$ is denoted by

$$\begin{aligned}h_{kl}&=\boldsymbol{e}_{\mathrm{r}}^{\mathrm{H}}(k/L_{\mathrm{r}})\left(\sum_{i=1}^{P}a_p\boldsymbol{e}_{\mathrm{r}}(\Omega_{tp})\boldsymbol{e}_{\mathrm{t}}^{\mathrm{H}}(\Omega_{tp})\right)\boldsymbol{e}_{\mathrm{t}}(l/L_{\mathrm{t}})\\&=\sum_{p=1}^{P}a_p\boldsymbol{e}_{\mathrm{r}}^{\mathrm{H}}(k/L_{\mathrm{r}})\boldsymbol{e}_{\mathrm{r}}(\Omega_{rp})\boldsymbol{e}_{\mathrm{t}}^{\mathrm{H}}(\Omega_{tp})\boldsymbol{e}_{\mathrm{t}}(l/L_{\mathrm{t}})\\&=\sum_{p=1}^{P}a_pf_{\mathrm{r}}(k/L_{\mathrm{r}}-\Omega_{rp})f_{\mathrm{t}}(\Omega_{tp}-l/L_{\mathrm{t}}).\end{aligned}\quad(9)$$

According to Eq. (5), when the direction of paths is an integer multiple of resolution $1/L_{\mathrm{r}}$, $1/L_{\mathrm{t}}$, the angular correlation f_{r} has the property of the Dirichlet function. There is no leakage in the channel matrix, which indicates that the entry h_{kl} represents the channel gain transmitted from the lth angular direction to the kth angular direction (Tse and Viswanath, 2005). For better legibility, the nth row of $\boldsymbol{H}_i^{\mathrm{a}}$ is denoted by $\boldsymbol{h}_{in}$:

$$\boldsymbol{h}_{in}=[h_{in}(1),h_{in}(2),\cdots,h_{in}(M)],1\le i<K,1\le n\le N.\quad(10)$$

Denote $D_{in}=\mathrm{supp}(\boldsymbol{h}_{in})$ as the support of $\boldsymbol{h}_{in}$. Extensive experimental studies (Barbotin *et al*., 2012; Hoydis *et al*., 2012) have shown that the channel matrix $\boldsymbol{H}_i^{\mathrm{a}}$ tends to be sparse because there are only a limited number of significant scatterers in the wireless propagation environments, which is an important property to be fully exploited. Specifically, for user i,

each row vector of $\boldsymbol{H}_i^{\mathrm{a}}$ has the same sparsity support:

$$D_{i1}=D_{i2}=\cdots=D_{iN}=\Gamma_i, \tag{11}$$

where Γ_i denotes the index set of support for user i. Eq. (11) reveals that the channel matrix of each user has the same support due to the finite scatterers of the channel at the BS and abundant scattering around the user (Gao *et al*., 2011; Rao and Lau, 2014) (Fig. 2). On the other hand, due to limited scatterers surrounding the BS and the channel correlation of users close to each other (Rao and Lau, 2014), different users have the same partial support:

$$\bigcap_{i=1}^{K}\Gamma_i=\Gamma_{\mathrm{c}}, \tag{12}$$

where Γ_{c} denotes the index set of common support among all K users. Eq. (12) explains that the channel matrices of different users are partially correlated due to some common scatterers (Gao *et al*., 2011) (Fig. 2). Furthermore, the channel sparsity support of multiple users can be parameterized as $\Gamma=\{\Gamma_{\mathrm{c}},\ \{\Gamma_i:\ \forall i\},\ 1\le i\le K\}$. The channel property in the angular domain above is summarized in the following definition:

Definition 1 (Statistical channel model) (Rao and Lau, 2014) For a massive MIMO system with K users, the channel sparsity support $\Gamma=\{\Gamma_{\mathrm{c}},\ \{\Gamma_i:\ \forall i\},\ 1\le i\le K\}$ refers to the statistical model $S=\{s_{\mathrm{c}},\ s_{\mathrm{p}}$: for K users$\}$, where $|\Gamma_{\mathrm{c}}|\ge s_{\mathrm{c}}$, $|\Gamma_i|\le s_{\mathrm{p}}$, and s_{c} and s_{p} denote the common support set and individual support set, respectively. In particular, BS has the knowledge of S.

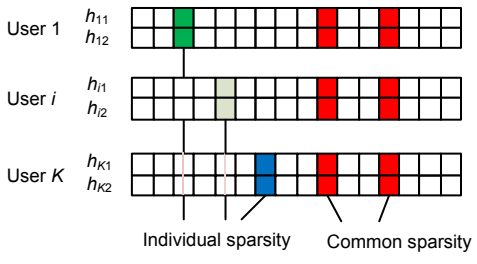

Fig. 2 Channel sparsity model

The statistical channel model relies on the practical scattering environment and remains invariable in a certain time slot. Hence, knowledge of the statistical model $S=\{s_{\mathrm{c}},\ s_{\mathrm{p}}$: for K users$\}$ can be known as prior information for the BS by offline channel propagation measurement (Gao *et al*., 2011; Rao and Lau, 2014). Note that it is difficult for the BS and the user i to obtain the instantaneous sparsity levels Γ_{c} and Γ_i due to the statistics of the wireless channel. Fortunately, the statistical model S makes channel estimation executable and possible in a practical system.

Remark 1 In this statistical channel model, when $s_{\mathrm{c}}=s_{\mathrm{p}}$, the model reduces to all users sharing the same local scatterers at the BS side. Also, when $s_{\mathrm{c}}=0$, the scenario becomes that the users share no common scatterers.

Remark 2 When each user is equipped with only one antenna, Eq. (11) degenerates to $D_{i1}=\Gamma_i$. That is to say, the same sparsity support among different antennas for user i disappears and the channel estimation performance worsens. The concrete analysis will be given in Section 3.

3 Proposed structured joint channel estimation algorithm

3.1 Multi-user massive MIMO model

In a typical multi-user massive MIMO system (Fig. 1), the BS equipped with M antennas simultaneously serves K users with $M\gg K$. The information transmission between the BS and each user is considered within the coherence time through a block fading channel, in which the channel is assumed to remain static for the duration of the coherence time. A specific communication process between the BS and users includes three stages: pilot sequence broadcasting, CSI feedback, and data transmission. Consider the pilot training stage with total time slots T and the transmitted pilot in tth (t=1, 2, ..., T) time slot denoted by $\boldsymbol{x}_t\in\mathbb{C}^{M\times 1}$. For the ith user, the $N\times 1$ received signal vector $\boldsymbol{y}_{it}$ is given by

$$\boldsymbol{y}_{it}=\boldsymbol{H}_i\boldsymbol{x}_t+\boldsymbol{n}_{it},\ t=1,2,\cdots,T, \tag{13}$$

where $\boldsymbol{H}_i\in\mathbb{C}^{N\times M}$ and $\boldsymbol{n}_{it}\in\mathbb{C}^{N\times 1}$ denote the channel matrix from the BS to the ith user and the additive white Gaussian noise vector, respectively. Considering all the T time slots, it is convenient to transfer Eq. (13) into matrix form as

$$\boldsymbol{Y}_i=\boldsymbol{H}_i\boldsymbol{X}+\boldsymbol{N}_i,\ 1\le i\le K, \tag{14}$$

where $\boldsymbol{X}=[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T] \in \mathbb{C}^{M\times T}$ is the transmitted pilot sequence in all T time slots, $\boldsymbol{Y}_i=[\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, \cdots, \boldsymbol{y}_{iT}] \in \mathbb{C}^{N\times T}$ and $\boldsymbol{N}_i=[\boldsymbol{n}_{i1}, \boldsymbol{n}_{i2}, \cdots, \boldsymbol{n}_{iT}] \in \mathbb{C}^{N\times T}$ are the received signal and noise matrices of the ith user in T time slots, respectively. Define P as the transmit signal-to-noise ratio (SNR) in each time slot at the BS. For brevity, the pilot sequences are supposed to have equal energy.

Because CSI is important for exploiting the DoF and array gain, the transceiver should acquire accurate CSI. For a massive MIMO system working in the FDD protocol, CSI acquisition at the BS usually consists of two steps: (1) local users estimate the CSI by using the pilot sequences; (2) the BS obtains the estimated CSI feedback via the reverse links. Users can employ the conventional channel estimation algorithm to obtain the CSI. In particular, the LS channel estimation $\tilde{\boldsymbol{H}}_i$ of the ith user can be expressed by

$$\tilde{\boldsymbol{H}}_i=\boldsymbol{Y}_i\boldsymbol{X}^{\dagger}, \tag{15}$$

$$\boldsymbol{X}^{\dagger}=\boldsymbol{X}^{\mathrm{H}}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}})^{-1}, \tag{16}$$

where $\boldsymbol{X}^{\dagger}$ is the Moore-Penrose inversion of $\boldsymbol{X}$. Note that when applying the LS-based algorithm, Eqs. (15) and (16) indicate that $T\geq M$, which signifies that a massive MIMO system requires a large amount of pilot training overhead and CSI feedback overhead due to the large number of BS antennas. Therefore, the conventional LS-based channel estimation algorithm is unsuitable for a massive MIMO scenario and it is imperative to design a more efficient scheme. Fortunately, by exploiting the sparsity in the angular domain of the channel matrix, the CS-based channel estimator can be applied to address this challenging problem. Furthermore, instead of the instantaneous sparsity levels Γ_{c} and Γ_i of wireless channels, the statistical channel model in Definition 1 can be fully exploited in a massive MIMO system to effectively reduce the pilot overhead.

3.2 Proposed SJMP algorithm

To fully exploit the joint channel sparsity property mentioned above, we first reformulate Eq. (14) by taking Eq. (8) into consideration, and then obtain

$$\boldsymbol{Y}_i'=\boldsymbol{H}_i^{\mathrm{a}}\boldsymbol{X}'+\boldsymbol{N}_i',\ 1\leq i\leq K, \tag{17}$$

where $\boldsymbol{Y}_i'=\boldsymbol{\Theta}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{Y}_i \in \mathbb{C}^{N\times T}$, $\boldsymbol{X}'=\boldsymbol{\Theta}_{\mathrm{T}}^{\mathrm{H}}\boldsymbol{X} \in \mathbb{C}^{M\times T}$, and $\boldsymbol{N}_i'=\boldsymbol{\Theta}_{\mathrm{R}}^{\mathrm{H}}\boldsymbol{N}_i \in \mathbb{C}^{N\times T}$.

Note that although the angular channel matrix $\boldsymbol{H}_i^{\mathrm{a}}$ tends to be sparse, Eq. (17) is not a typical CS problem. It is very challenging to solve Eq. (17) using traditional CS-based algorithms. Furthermore, the complexity is high due to the large dimension of $\boldsymbol{X}'=\mathbb{C}^{M\times T}$ in a massive MIMO system.

Fortunately, the channel structure can be fully exploited. We reformulate Eq. (17) using the matrix vector operator as

$$\begin{aligned}\mathrm{vec}(\boldsymbol{Y}_i')&=\mathrm{vec}(\boldsymbol{H}_i^{\mathrm{a}}\boldsymbol{X}'+\boldsymbol{N}_i')=\mathrm{vec}(\boldsymbol{H}_i^{\mathrm{a}}\boldsymbol{X}')+\mathrm{vec}(\boldsymbol{N}_i')\\&=\boldsymbol{X}_{\mathrm{B}}'\cdot\mathrm{vec}(\boldsymbol{H}_i^{\mathrm{a}})+\mathrm{vec}(\boldsymbol{N}_i'),\end{aligned} \tag{18}$$

where $\boldsymbol{X}_{\mathrm{B}}'=\boldsymbol{X}'^{\mathrm{T}}\otimes\boldsymbol{I}_N$. Note that $\mathrm{vec}(\boldsymbol{H}_i^{\mathrm{a}})$ is a sparse vector and $\boldsymbol{X}'^{\mathrm{T}}\otimes\boldsymbol{I}_N$ is the measurement matrix, which indicates that Eq. (18) is a typical CS formula. In the frame of CS, to guarantee the recovery performance of CS reconstruction, a relatively small mutual coherence of the measurement matrix is always required. Note that $\boldsymbol{X}'^{\mathrm{T}}$ is composed by pilot matrix $\boldsymbol{X}$, which experiences unitary matrix multiplication and transpose operation. The mutual coherence of $\boldsymbol{X}'^{\mathrm{T}}$ is defined as (Qi and Wu, 2014)

$$\mu(\boldsymbol{X}'^{\mathrm{T}})=\max_{\substack{1\leq i,l\leq M\\ i\neq l}}\frac{\left|\left\langle\boldsymbol{x}_i',\boldsymbol{x}_l'\right\rangle\right|}{\left\|\boldsymbol{x}_i'\right\|_2\left\|\boldsymbol{x}_l'\right\|_2}, \tag{19}$$

where $\boldsymbol{x}_i'$ and $\boldsymbol{x}_l'$ denote the ith and lth columns of matrix $\boldsymbol{X}'^{\mathrm{T}}$, respectively. According to Eq. (11), when the user is equipped with more than one antenna, the support of the channel matrix among different user antennas is identical. This leads to a 'block' property in $\mathrm{vec}(\boldsymbol{H}_i^{\mathrm{a}})$. The 'block' property means that the consecutive N elements in channel vector $\mathrm{vec}(\boldsymbol{H}_i^{\mathrm{a}})$ are either all non-zero or all zero. Consequently, the block coherence of $\boldsymbol{X}_{\mathrm{B}}'$ is defined as (Eldar *et al*., 2010)

$$\mu_{\mathrm{B}}(\boldsymbol{X}'_{\mathrm{B}})=\frac{1}{N}\max_{\substack{1\le i,l\le M\\ i\ne l}}\rho\left(\boldsymbol{x}'^{\mathrm{H}}_{\mathrm{B}i}\boldsymbol{x}'_{\mathrm{B}l}\right), \tag{20}$$

where $\boldsymbol{x}'_{\mathrm{B}i}$ and $\boldsymbol{x}'_{\mathrm{B}l}$ denote the ith and lth column blocks of matrix $\boldsymbol{X}'_{\mathrm{B}}$, respectively. The spectrum radius of a given matrix $\boldsymbol{A}$ is denoted as

$$\rho(\boldsymbol{A})=\max_i\sqrt{\lambda_i(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}, \tag{21}$$

where $\lambda_i(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$ denotes the ith eigenvalue of matrix $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$. The relationship between Eqs. (19) and (20) is given in the following theorem:

Theorem 1 For the channel estimation formula in Eq. (17), we have

$$\mu_{\mathrm{B}}(\boldsymbol{X}'_{\mathrm{B}})=\frac{1}{N}\mu(\boldsymbol{X}'^{\mathrm{T}}). \tag{22}$$

Proof See the Appendix.

Theorem 1 signifies that the block coherence of the measurement matrix is no more than the mutual coherence for the pilot sequence matrix. In the field of CS, a smaller coherence guarantees a better recovery performance of the original sparse signal. When the user has a single antenna, the 'block' coherence is identical to the mutual coherence, which is the case considered in most previous studies. The channel estimation performance cannot be improved. When N>1, the block coherence reduces and the column blocks tend to be orthogonal with each other, which leads to better channel recovery. Consequently, the pilot overhead can be further reduced due to the block coherence while holding the same channel performance.

Furthermore, through conjugate transpose operation, Eq. (17) can be denoted as

$$\bar{\boldsymbol{Y}}_i=\bar{\boldsymbol{X}}\bar{\boldsymbol{H}}_i+\bar{\boldsymbol{N}}_i,\ 1\le i\le K, \tag{23}$$

where $\bar{\boldsymbol{Y}}_i=\boldsymbol{Y}_i^{\mathrm{H}}\boldsymbol{\Theta}_{\mathrm{R}}\in\mathbb{C}^{T\times N}$, $\bar{\boldsymbol{X}}=\boldsymbol{X}^{\mathrm{H}}\boldsymbol{\Theta}_{\mathrm{T}}\in\mathbb{C}^{T\times M}$, $\bar{\boldsymbol{H}}_i=(\boldsymbol{H}_i^{\mathrm{a}})^{\mathrm{H}}\in\mathbb{C}^{M\times N}$, and $\bar{\boldsymbol{N}}_i=\boldsymbol{N}_i^{\mathrm{H}}\boldsymbol{\Theta}_{\mathrm{R}}\in\mathbb{C}^{T\times N}$.

Note that Eq. (23) is a standard multiple measurement vector (MMV) problem in CS models. Popular greedy algorithms, SOMP algorithms, and the subsequent algorithm based on SOMP are proposed to solve MMV problems. Further, when the number of antennas N on the user side reduces to 1, problem (23) degenerates to a standard single measurement vector (SMV) problem, which can be solved using conventional OMP (Tropp and Gilbert, 2007) and SP (Dai and Milenkovic, 2009) algorithms. As Theorem 1 shows, the block coherence of the pilot sequence matrix is equal to the mutual coherence. However, these algorithms do not consider the common sparsity structures of channel matrices among different users. Therefore, it is not suitable to address problem (23) using the associated algorithms when the user has more than one antenna.

The communication process between the BS and users is illustrated below. First, the BS broadcasts the pilot sequences $\boldsymbol{X}\in\mathbb{C}^{M\times T}$ with $T\ll M$ to all the K users. Then, user i obtains the channel measurements $\boldsymbol{Y}_i$ according to problem (23) and then feeds $\boldsymbol{Y}_i$ back to the BS. Finally, the BS obtains the channel measurements $[\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_K]$ of all K users and jointly processes and obtains the estimated channel matrix $[\boldsymbol{H}_1^{\mathrm{e}}, \boldsymbol{H}_2^{\mathrm{e}}, \cdots, \boldsymbol{H}_K^{\mathrm{e}}]$. In particular, the channel recovery on the BS side is formulated as

$$\begin{aligned}&\bar{\boldsymbol{H}}_i^{\mathrm{e}}=\min_{\{\bar{\boldsymbol{H}}_i\}}\sum_{i=1}^{K}\left\|\bar{\boldsymbol{Y}}_i-\bar{\boldsymbol{X}}\bar{\boldsymbol{H}}_i\right\|_{\mathrm{F}}^2,\ 1\le i\le K,\\ &\text{s.t. }\{\bar{\boldsymbol{H}}_i,\ 1\le i\le K\}\text{ satisfies the sparsity property.}\end{aligned} \tag{24}$$

To fully leverage the common sparsity of channel matrices in multi-user massive MIMO systems, the proposed algorithm is elaborated in the next subsection. In this subsection, the SJMP algorithm, Algorithm 1, is proposed to fully leverage the specific sparsity among the channel matrices of the multi-user massive MIMO system mentioned in Definition 1.

Note that the row elements of $\bar{\boldsymbol{H}}_i^{\mathrm{e}}$ are either all zero or all nonzero, which is similar to the case of MMV problems. Developed from the SOMP algorithm, the proposed SJMP algorithm exploits the structured sparsity of channel matrices to improve channel estimation recovery performance. Here, we further explain the SJMP. After the initialization, steps 2–4 aim to identify the common support of channel matrices among all the users. Next, set the initial individual support with the estimated common

support for each user, and then follow steps 6–9 to obtain the channel support for each user. Finally, based on the solution of step 9, we acquire the final estimated channels.

Algorithm 1 Structured joint matching pursuit

Input: $\{\boldsymbol{Y}_i, 1\le i\le K\}$, $\boldsymbol{X}$, $S=\{s_\mathrm{c}, s_\mathrm{p}$: for K users$\}$.

1: (Initialization) Compute $\{\bar{\boldsymbol{Y}}_i, 1\le i\le K\}$ and $\bar{\boldsymbol{X}}$ according to Eq. (23); the initial set $\Gamma_\mathrm{c}^\mathrm{e}=\varphi$, $\Gamma_i'=\varphi$, and the residue $\{\boldsymbol{R}_i=\bar{\boldsymbol{Y}}_i, 1\le i\le K\}$.

Iteration 1: Repeat the following steps s_c times.

2: (Support estimate) $\Gamma_i'=\Gamma_i'\cup\{$indices corresponding to the s_p largest l_2-norm of the row vectors of $\bar{\boldsymbol{X}}^\mathrm{H}\boldsymbol{R}_i\}$;

3: (Support pruning):

$$\Gamma_\mathrm{c}^\mathrm{e}=\Gamma_\mathrm{c}^\mathrm{e}\cup\left\{\arg\max_j\sum_{i=1}^{K}I_{\{j\in\Gamma_i'\}}\right\};$$

4: (Residue update):

$$\boldsymbol{R}_i=\bar{\boldsymbol{Y}}_i-\bar{\boldsymbol{X}}_{\downarrow\Gamma_\mathrm{c}^\mathrm{e}}(\bar{\boldsymbol{X}}_{\downarrow\Gamma_\mathrm{c}^\mathrm{e}})^\dagger\bar{\boldsymbol{Y}}_i,\quad \Gamma_i'=\varphi;$$

End iteration 1

5: (Support initialization):

Set $\Gamma_i^\mathrm{e}=\Gamma_\mathrm{c}^\mathrm{e}$ for $1\le i\le K$;

Iteration 2: Go through the following steps $s_\mathrm{p}-s_\mathrm{c}$ times for each user.

6: (Support update) $\Gamma_i^\mathrm{e}=\Gamma_i^\mathrm{e}\cup\{$indices corresponding to the s_p largest l_2-norm of the row vectors of $\bar{\boldsymbol{X}}^\mathrm{H}\boldsymbol{R}_i\}$;

7: (Projection) $\tilde{\boldsymbol{X}}=\bar{\boldsymbol{X}}_{\downarrow\Gamma_i}^\dagger\boldsymbol{Y}_i$;

8: (Individual support estimate) $\Gamma_i^\mathrm{e}=\{$indices corresponding to the s_p largest l_2-norm of the row vectors of $\tilde{\boldsymbol{X}}\}$;

9: (Residue update) $\boldsymbol{R}_i=\bar{\boldsymbol{Y}}_i-\bar{\boldsymbol{X}}_{\downarrow\Gamma_i^\mathrm{e}}(\bar{\boldsymbol{X}}_{\downarrow\Gamma_i^\mathrm{e}})^\dagger\bar{\boldsymbol{Y}}_i$;

End iteration 2

10: (Obtaining CSI) The estimated support for each user satisfies $(\bar{\boldsymbol{H}}_i^\mathrm{e})^{\{\{1,2,\cdots,M\}-\Gamma_i^\mathrm{e}\}\rightarrow}=\boldsymbol{0}$ and $(\bar{\boldsymbol{H}}_i^\mathrm{e})^{\Gamma_i^\mathrm{e}\rightarrow}=(\bar{\boldsymbol{X}}_{\downarrow\Gamma_i^\mathrm{e}})^\dagger\bar{\boldsymbol{Y}}_i$; the channel matrix can be obtained according to Eq. (23).

Output: $[\boldsymbol{H}_1^\mathrm{e}, \boldsymbol{H}_2^\mathrm{e}, \cdots, \boldsymbol{H}_K^\mathrm{e}]$.

Compared with SOMP, the proposed algorithm exploits the common sparsity structure of the multi-user massive MIMO system. Instead of recovering the support set directly for each user, the proposed SJMP firstly identifies the support index which is likely to belong to all the users as the common support set. Then, on the basis of common support, the proposed algorithm iteratively recovers the individual support. It fully exploits the statistical channel structure as in Definition 1. Moreover, the input parameters are another highlight. Unlike conventional CS-based sparse channel estimation (Hu *et al*., 2013; Qi and Wu, 2014) and algorithms (Rao and Lau, 2014; Chen and Qin, 2015), where accurate and instantaneous sparsity is needed, the proposed algorithm demands only the statistical channel sparsity on the BS side, which can be easily obtained by channel measurements. In addition, the proposed algorithm has no need for extra empirical parameters (Rao and Lau, 2014), which enables SJMP to be more flexible and easier to operate. Furthermore, the proposed algorithm incorporates the merit of the SP algorithm, where the support set is refined by projection in step 7, so that the channel estimation quality can be improved further compared with J-OMP. In fact, the sparsity can be expressed in two categories: instantaneous sparsity and statistical channel sparsity. Instantaneous sparsity can vary within a certain range of statistical channel sparsity. Since it is difficult to know the exact instantaneous sparsity level in practice, the proposed algorithm fully exploits the statistical channel sparsity as a priority to acquire the channel matrix of different users and cannot acquire the sparsity level adaptively. At the expense of not adaptively acquiring the instantaneous sparsity level, the proposed algorithm can be more flexible and easier to operate in a practical massive MIMO system.

Remark 3 When $s_\mathrm{c}=s_\mathrm{p}$, all users share the same common support. Steps 6 and 7 in SJMP are skipped and the proposed algorithm degenerates to the standard SOMP. Also, when $s_\mathrm{c}=0$, there is no common support shared by all the users. Steps 2–4 in SJMP are skipped and the proposed algorithm also degenerates to the standard SOMP. Hence, the conventional SOMP can be considered as a special case of our proposed SJMP algorithm.

4 Performance analysis

4.1 Probability of recovering common support

To implement analysis of SJMP, the restricted isometry property (RIP) for our problem (23) will be illustrated first. Given a k-order restricted isometry constant (RIC) δ_k ($\delta_k\in[0, 1)$) and an arbitrary vector $\boldsymbol{h}\in\mathbb{C}^{M\times1}$ ($\|\boldsymbol{h}\|_0\le k$), matrix $\bar{\boldsymbol{X}}\in\mathbb{C}^{T\times M}$ in problem (23) satisfies RIP when holding

$$(1-\delta_k)\|\boldsymbol{h}\|^2\le\|\bar{\boldsymbol{X}}\boldsymbol{h}\|^2\le(1+\delta_k)\|\boldsymbol{h}\|^2. \quad (25)$$

Lemma 1 (Dai and Milenkovic, 2009) For two disjoint sets J_1, $J_2 \subset \{1, 2, \ldots, M\}$, assuming $\delta_{|J_1|+|J_2|} < 1$, then for arbitrary vectors $\boldsymbol{h}_1 \in \mathbb{R}^{|J_1|\times 1}$ and $\boldsymbol{h}_2 \in \mathbb{R}^{|J_2|\times 1}$, we obtain

$$\left|\left\langle \bar{\boldsymbol{X}}_{\downarrow J_1}\boldsymbol{h}_1, \bar{\boldsymbol{X}}_{\downarrow J_2}\boldsymbol{h}_2\right\rangle\right| \le \delta_{|J_1|+|J_2|}\left\|\boldsymbol{h}_1\right\|_2\left\|\boldsymbol{h}_2\right\|_2, \tag{26}$$

$$\left\|\bar{\boldsymbol{X}}_{\downarrow J_2}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J_2}\boldsymbol{h}_2\right\|_2 \le \delta_{|J_2|}\left\|\boldsymbol{h}_2\right\|_2. \tag{27}$$

Proof Define $\boldsymbol{h}_1' = \boldsymbol{h}_1/\|\boldsymbol{h}_1\|$, $\boldsymbol{x}' = \bar{\boldsymbol{X}}_{\downarrow J_1}\boldsymbol{h}_1'$, $\boldsymbol{h}_2' = \boldsymbol{h}_2/\|\boldsymbol{h}_2\|$, and $\boldsymbol{y}' = \bar{\boldsymbol{X}}_{\downarrow J_2}\boldsymbol{h}_2'$. Then, we have

$$\begin{aligned}&\frac{\left|\left\langle \bar{\boldsymbol{X}}_{\downarrow J_1}\boldsymbol{h}_1, \bar{\boldsymbol{X}}_{\downarrow J_2}\boldsymbol{h}_2\right\rangle\right|}{\left\|\boldsymbol{h}_1\right\|_2\left\|\boldsymbol{h}_2\right\|_2}\\ &= \left|\left\langle \bar{\boldsymbol{X}}_{\downarrow J_1}\boldsymbol{h}_1', \bar{\boldsymbol{X}}_{\downarrow J_2}\boldsymbol{h}_2'\right\rangle\right| = \left|\left\langle \boldsymbol{x}', \boldsymbol{y}'\right\rangle\right|\\ &= \frac{\left\|\boldsymbol{x}'+\boldsymbol{y}'\right\|_2^2 - \left\|\boldsymbol{x}'-\boldsymbol{y}'\right\|_2^2}{4} \le \delta_{|J_1|+|J_2|}.\end{aligned} \tag{28}$$

Thus, inequality (26) is proved. Now, we have

$$\begin{aligned}\left\|\bar{\boldsymbol{X}}_{\downarrow J_2}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J_2}\boldsymbol{h}_2\right\|_2 &= \max_{\boldsymbol{p}:\|\boldsymbol{p}\|_2=1}\left|\boldsymbol{p}^{\mathrm{H}}(\bar{\boldsymbol{X}}_{\downarrow J_2}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J_2}\boldsymbol{h}_2)\right|\\ &\le \max_{\boldsymbol{p}:\|\boldsymbol{p}\|_2=1}\delta_{|J_1|+|J_2|}\left\|\boldsymbol{p}\right\|_2\left\|\boldsymbol{h}_2\right\|_2\\ &= \delta_{|J_1|+|J_2|}\left\|\boldsymbol{h}_2\right\|_2.\end{aligned} \tag{29}$$

The proof of Lemma 1 is completed.

This lemma explains the relationship between pilot sequences and sparse channel support, which will be used frequently.

Lemma 2 (Orthogonality) For a set J and pilot sequence matrix $\bar{\boldsymbol{X}}$, denote $\bar{\boldsymbol{X}}_{\downarrow J}$ as the sub-matrix consisting of columns of $\bar{\boldsymbol{X}}$ whose indices belong to J. Then

$$\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}(\boldsymbol{I}-\boldsymbol{P}_J) = \boldsymbol{0}, \tag{30}$$

where $\boldsymbol{I}$ is the identity matrix and

$$\boldsymbol{P}_J = \bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}. \tag{31}$$

Proof

$$\begin{aligned}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\boldsymbol{R}_i &= \bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\left[\bar{\boldsymbol{Y}}_i - \bar{\boldsymbol{X}}_{\downarrow J}\left(\bar{\boldsymbol{X}}_{\downarrow J}\right)^{\dagger}\bar{\boldsymbol{Y}}_i\right]\\ &= \bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\left[\boldsymbol{I} - \bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\right]\bar{\boldsymbol{Y}}_i\\ &= \bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\left[\bar{\boldsymbol{Y}}_i - \bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{Y}}_i\right]\\ &= \bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{Y}}_i - \bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{Y}}_i\\ &= \bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{Y}}_i - \bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{Y}}_i\\ &= \boldsymbol{0}.\end{aligned} \tag{32}$$

Note that residual $\bar{\boldsymbol{Y}}_i \neq \boldsymbol{0}$. We obtain

$$\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}(\boldsymbol{I} - \bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}) = \boldsymbol{0}, \tag{33}$$

which completes the proof.

Lemma 2 indicates that the sub-matrix containing columns corresponding to support set J of $\bar{\boldsymbol{X}}$ is orthogonal with the residual matrix in the SJMP algorithm. This property will be used frequently in the following proof.

Now we take a close look at the process of common support identification in SJMP. Common support identification is closely related to the whole recovery performance. We use $\Pr(E_{\mathrm{C}})$ to represent the probability of the event of correctly identifying common support. Intuitively, the higher the probability of common support recovery $\Pr(E_{\mathrm{C}})$, the more accurate the channel estimation $\boldsymbol{H}_i^{\mathrm{e}}$ is, Specifically, by using the above Lemmas 1 and 2, we can obtain the probability bound of $\Pr(E_{\mathrm{C}})$ for our proposed SJMP algorithm.

Theorem 2 (Upper bound of $\Pr(E_{\mathrm{C}})$) For parameter β, if

$$\beta = \frac{(1-\delta_{s_{\mathrm{p}}}-\delta_{s_{\mathrm{c}}})^3}{(1-\delta_{s_{\mathrm{c}}})\left(\delta_{s_{\mathrm{p}}+1}+\delta_{s_{\mathrm{p}}+1}\delta_{s_{\mathrm{p}}}+2\sqrt{\dfrac{(1-\delta_{s_{\mathrm{c}}})^2(1+\delta_1)T}{P}}\right)^2} > 1 \tag{34}$$

holds, then the probability of the event of correctly identifying common support $\Pr(E_{\mathrm{C}})$ satisfies

$$\Pr(E_{\mathrm{C}}) \geq 1-\sum_{t=0}^{K_1}\binom{K}{t}(1-p)^t p^{K-t}-\sum_{t=0}^{K-K_1}\binom{K}{t}(1-p)^t p^{K-t}, \tag{35}$$

where δ_1, δ_{s_c}, δ_{s_p}, and δ_{s_p+1} are the 1st, s_cth, s_pth, and (s_p+1)th RIC of $\bar{\boldsymbol{X}}$, respectively, and

$$\begin{aligned} p=&\exp\left(-N\left(\frac{1-\beta^2}{\beta^2}+2\ln\beta\right)\right)\\ &+\exp\left(-N\left[\frac{(2-\delta_{s_p}-\delta_{s_c})\delta_{s_p}}{(1-\delta_{s_p}-\delta_{s_c})^2}+2\ln\frac{1-\delta_{s_c}}{1-\delta_{s_p}-\delta_{s_c}}\right]\right)\\ &+\exp(-N(2T-1-\ln 2T)). \end{aligned} \tag{36}$$

Proof Note that if the event correctly identifying common support occurs, it means the added index in step 2 belongs to the correct support set Γ_c. For better comprehension, denote J as the current estimated common support. Then the next estimated set $\Gamma_i'(J)$, according to steps 2–4, can be obtained by

$$\Gamma_i'=\arg\max_{|\Gamma|=s_p}\left\|\bar{\boldsymbol{X}}_{\downarrow\Gamma}^{\mathrm{H}}(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{Y}}_i\right\|_{\mathrm{F}}. \tag{37}$$

Next, for the current estimated common support J and any $l\in J$, $l\notin\Gamma_i'(J)$, we have

$$\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{Y}}_i=\boldsymbol{0},\ \forall l\in J, \tag{38}$$

$$\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{Y}}_i\neq\boldsymbol{0},\ \forall l\in[1,2,\cdots,M]\setminus J. \tag{39}$$

The equation comes from steps 2–4 in SJMP, which indicates that any support atom obtained in the previous iteration will not be picked again in the next iteration, ensuring the uniqueness of atom selection in each iteration. Furthermore, given the current estimated common support as $J\subseteq\Gamma_c$, $|J|<s_c$, the index l obtained in the next iteration will belong to Γ_c when the following event occurs:

$$\varepsilon_J:\ \max_{l\in\Gamma_c\setminus J}\sum_{i=1}^{K}I_{\{l\in\Gamma_i'(J)\}}>\max_{j\in[1,2,\cdots,M]\setminus\Gamma_c}\sum_{i=1}^{K}I_{\{j\in\Gamma_i'(J)\}}. \tag{40}$$

Therefore, to guarantee that the event of correctly identifying common support occurs, the above equation should apply for all $J\subseteq\Gamma_c$, $|J|<s_c$. If we cannot identify the common support correctly for all $J\subseteq\Gamma_c$, $|J|<s_c$, Eq. (40) will not hold. Then we have

$$\Pr(\bar{E}_{\mathrm{C}})\leq\Pr\Big(\bigcup_{J\subseteq\Gamma_c,|J|<s_c}\bar{\varepsilon}_J\Big)\leq\bigcup_{J\subseteq\Gamma_c,|J|<s_c}\Pr(\bar{\varepsilon}_J). \tag{41}$$

For the current estimated common support $J\subseteq\Gamma_c$, $|J|<s_c$, and $l\in\Gamma_c\setminus J$, we have

$$\begin{aligned} \Pr(\bar{\varepsilon}_J)=&\Pr\left(\max_{l\in\Gamma_c\setminus J}\sum_{i=1}^{K}I_{\{l\in\Gamma_i'(J)\}}\leq\max_{j\in[1,2,\cdots,M]\setminus\Gamma_c}\sum_{i=1}^{K}I_{\{j\in\Gamma_i'(J)\}}\right)\\ \leq&\Pr\left(\sum_{i=1}^{K}I_{\{l\in\Gamma_i'(J)\}}\leq K_1\right)\\ &+\Pr\left(\max_{j\in[1,2,\cdots,M]\setminus\Gamma_c}\sum_{i=1}^{K}I_{\{j\in\Gamma_i'(J)\}}\geq K_1\right). \end{aligned} \tag{42}$$

The inequality is the property of probability and K_1 is a constant indicating the number of users, which determines whether a support index belongs to the common sparsity. For any $j\in[1, 2, \ldots, M]\setminus\Gamma_c$, we have

$$\begin{aligned} &\Pr\left(\max_{j\in[1,2,\cdots,M]\setminus\Gamma_c}\sum_{i=1}^{K}I_{\{j\in\Gamma_i'(J)\}}\geq K_1\right)\\ &\leq\Pr\left(\sum_{i=1}^{K}\max_{j\in[1,2,\cdots,M]\setminus\Gamma_c}I_{\{j\in\Gamma_i'(J)\}}\geq K_1\right). \end{aligned} \tag{43}$$

According to step 2 in Algorithm 1, for a given current common support set $J\subseteq\Gamma_c$, $|J|<s_c$, and $l\in\Gamma_c\setminus J$, it holds that

$$\begin{aligned} \Pr(I_{\{l\in\Gamma_i'(J)\}}=0)=&\Pr\Big(\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{Y}}_i\right\|_{\mathrm{F}}^2\\ &\leq\max_{j\in[1,2,\cdots,M]\setminus\Gamma_c}\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{Y}}_i\right\|_{\mathrm{F}}^2\Big). \end{aligned} \tag{44}$$

Note that

$$\bar{\boldsymbol{Y}}_i=\bar{\boldsymbol{X}}\bar{\boldsymbol{H}}_i+\bar{\boldsymbol{N}}_i=\bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{H}}_i)^{J\rightarrow}+\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}+\bar{\boldsymbol{N}}_i. \tag{45}$$

Then, for the left side of the inequality, according to Lemmas 1 and 2, we have

$$\begin{aligned}
&\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{Y}}_i\right\|_{\mathrm{F}}\\
&=\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)(\bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{H}}_i)^{J\rightarrow}+\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}+\bar{\boldsymbol{N}}_i)\right\|_{\mathrm{F}}\\
&\overset{(a)}{=}\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)(\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}+\bar{\boldsymbol{N}}_i)\right\|_{\mathrm{F}}\\
&\overset{(b)}{\geq}\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}\right\|_{\mathrm{F}}-\sqrt{1+\delta_1}\left\|\bar{\boldsymbol{N}}_i\right\|_{\mathrm{F}},
\end{aligned}\tag{46}$$

where (a) follows Lemma 2 and (b) follows the triangle inequality and Eq. (25). On the other hand, for the right side of the inequality, we obtain

$$\begin{aligned}
&\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{Y}}_i\right\|_{F}\\
&=\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)(\bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{H}}_i)^{J\rightarrow}+\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}+\bar{\boldsymbol{N}}_i)\right\|_{\mathrm{F}}\\
&=\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)(\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}+\bar{\boldsymbol{N}}_i)\right\|_{\mathrm{F}}\\
&\leq\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}\right\|_{\mathrm{F}}+\sqrt{1+\delta_1}\left\|\bar{\boldsymbol{N}}_i\right\|_{\mathrm{F}}.
\end{aligned}\tag{47}$$

Combining Eqs. (46) and (47), we have

$$\begin{aligned}
&\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}\right\|_{\mathrm{F}}\\
&\leq\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}\right\|_{\mathrm{F}}+2\sqrt{1+\delta_1}\left\|\bar{\boldsymbol{N}}_i\right\|_{\mathrm{F}}.
\end{aligned}\tag{48}$$

Then Eq. (44) can be expressed as

$$\begin{aligned}
&\Pr(I_{\{l\in\Gamma_i'(J)\}}=0)\\
&\leq\Pr\left(\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}\right\|_{\mathrm{F}}\leq\sqrt{N}\alpha\right)\\
&+\Pr\left(\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}\right\|_{\mathrm{F}}\right.\\
&\qquad\left.\geq\sqrt{N}\alpha-2\sqrt{(1+\delta_1)TN/P}\right)\\
&+\Pr\left(\left\|\bar{\boldsymbol{N}}_i\right\|_{F}\geq\sqrt{TN/P}\right),
\end{aligned}\tag{49}$$

where α is a constant which will be set later. Furthermore, according to Lemma 1 again, for part of Eq. (46), we obtain

$$\begin{aligned}
&\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}\right\|_{\mathrm{F}}\\
&\geq\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}\right\|_{\mathrm{F}}-\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)\bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}\right\|_{\mathrm{F}}\\
&\geq 1-\delta_{s_{\mathrm{p}}}-\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)\bar{\boldsymbol{X}}_{\downarrow J}\right\|_{\mathrm{F}}\left\|(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\right\|_{\mathrm{F}}\left\|\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}\right\|_{\mathrm{F}}\\
&\geq\frac{1-\delta_{s_{\mathrm{p}}}-\delta_{s_{\mathrm{c}}}}{1-\delta_{s_{\mathrm{c}}}}.
\end{aligned}\tag{50}$$

For part of Eq. (47), we obtain

$$\begin{aligned}
&\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}\right\|_{\mathrm{F}}\\
&\leq\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}\right\|_{\mathrm{F}}+\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)\bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}\right\|_{\mathrm{F}}\\
&\leq\delta_{s_{\mathrm{p}}+1}+\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)\bar{\boldsymbol{X}}_{\downarrow J}(\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow J})^{-1}\bar{\boldsymbol{X}}_{\downarrow J}^{\mathrm{H}}\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}\right\|_{\mathrm{F}}\\
&\leq\frac{\delta_{s_{\mathrm{p}}+1}+\delta_{s_{\mathrm{p}}+1}\delta_{s_{\mathrm{p}}}}{1-\delta_{s_{\mathrm{c}}}}.
\end{aligned}\tag{51}$$

Note that $\bar{\boldsymbol{H}}_i$ and $\bar{\boldsymbol{N}}_i$ are subject to the complex Gaussian distribution with mean 0 and variance 1, $1/P$, respectively. Denote χ_{2N} as a chi-squared distribution with $2N$ degrees of freedom. From Eqs. (49) and (50), we have

$$\begin{aligned}
&\Pr\left(\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}\right\|_{\mathrm{F}}\leq\sqrt{N}\alpha\right)\\
&\leq\Pr\left(\chi_{2N}\leq 2N\alpha^2\Big/\left(\frac{1-\delta_{s_{\mathrm{p}}}-\delta_{s_{\mathrm{c}}}}{1-\delta_{s_{\mathrm{c}}}}\right)^2\right).
\end{aligned}\tag{52}$$

From Eqs. (49) and (51), we have

$$\begin{aligned}
&\Pr\left(\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{X}}_{\downarrow\Gamma_i\setminus J}(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus J\rightarrow}\right\|_{\mathrm{F}}\right.\\
&\qquad\left.\geq\sqrt{N}\alpha-2\sqrt{(1+\delta_1)TN/P}\right)\\
&\leq\Pr\left(\chi_{2N}\geq 2N\left[\alpha-2\sqrt{(1+\delta_1)T/P}\right]^2\right.\\
&\qquad\left.\cdot\frac{1-\delta_{s_{\mathrm{c}}}}{\delta_{s_{\mathrm{p}}+1}+\delta_{s_{\mathrm{p}}+1}\delta_{s_{\mathrm{p}}}}\right).
\end{aligned}\tag{53}$$

From Eq. (49), we have

$$\Pr\left(\left\|\bar{\boldsymbol{N}}_i\right\|_{\mathrm{F}}\geq\sqrt{TN/P}\right)=\Pr(\chi_{2NT}\geq 2TN).\tag{54}$$

Next, combining Eqs. (52)–(54), we have

$$\begin{aligned}
\Pr(I_{\{l\in\Gamma_i'(J)\}}=0)&\leq\Pr\left(\chi_{2N}\leq\frac{2N\alpha^2}{\left(\frac{1-\delta_{s_{\mathrm{p}}}-\delta_{s_{\mathrm{c}}}}{1-\delta_{s_{\mathrm{c}}}}\right)^2}\right)\\
&+\Pr\left(\chi_{2N}\geq\frac{2N(\alpha-2\sqrt{(1+\delta_1)T/P})^2}{\left(\frac{\delta_{s_{\mathrm{p}}+1}+\delta_{s_{\mathrm{p}}+1}\delta_{s_{\mathrm{p}}}}{1-\delta_{s_{\mathrm{c}}}}\right)^2}\right)+\Pr(\chi_{2NT}\geq 2NT).
\end{aligned}\tag{55}$$

According to the Chernoff bounds (Dasgupta and Gupta, 2003) for chi-squared distribution χ_{2k}, we have the following bounds:

$$\Pr(\chi_{2k} \le 2xk) \le \exp(-k(-1+x-\ln x)),\ 0<x<1, \quad (56)$$

$$\Pr(\chi_{2k} \ge 2xk) \le \exp(-k(-1+x-\ln x)),\ x>1. \quad (57)$$

To use Eqs. (56) and (57), we simply choose

$$\alpha = \left[\frac{\delta_{s_p+1}+\delta_{s_p+1}\delta_{s_p}}{1-\delta_{s_c}} + 2\sqrt{\frac{(1+\delta_1)T}{P}}\right]\frac{1-\delta_{s_c}}{1-\delta_{s_p}-\delta_{s_c}}.$$

If $\beta = \dfrac{1-\delta_{s_p}-\delta_{s_c}}{(1-\delta_{s_c})\alpha^2} > 1$, then Eq. (55) can be expressed as

$$\begin{aligned}\Pr(I_{\{l\in\Gamma_i'(J)\}}=0) &\le \exp\left(-N\left(2\ln\beta+\frac{1}{\beta^2}-1\right)\right)\\ &+\exp\left(-N\left(2\ln\frac{1-\delta_{s_c}}{1-\delta_{s_p}-\delta_{s_c}}+\frac{\left(1-\delta_{s_c}\right)^2}{\left(1-\delta_{s_p}-\delta_{s_c}\right)^2}-1\right)\right)\\ &+\exp(-N(2T-\ln 2T-1)) = p.\end{aligned} \quad (58)$$

Then, the probability $\Pr\left(\max\limits_{j\in[1,2,\cdots,M]\setminus\Gamma_c} I_{\{j\in\Gamma_i'(J)\}}=1\right)$ can be derived. In fact,

$$\begin{aligned}&\Pr\left(\max_{j\in[1,2,\cdots,M]\setminus\Gamma_c} I_{\{j\in\Gamma_i'(J)\}}=1\right)\\ &=\Pr\left(\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(l)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{Y}}_i\right\|_{\mathrm{F}}^2\right.\\ &\qquad\left.\le \max_{j\in[1,2,\cdots,M]\setminus\Gamma_c}\left\|\bar{\boldsymbol{X}}^{\mathrm{H}}(j)(\boldsymbol{I}-\boldsymbol{P}_J)\bar{\boldsymbol{Y}}_i\right\|_{\mathrm{F}}^2\right)\\ &=\Pr(I_{\{l\in\Gamma_i'(J)\}}=0)\le p.\end{aligned} \quad (59)$$

$I_{\{l\in\Gamma_i'(J)\}}$ can be regarded as a statistics variable and is independent among different users because of the statistics of channel matrix $\bar{\boldsymbol{H}}_i$ for the ith user. Thus, $\left\{I_{\{l\in\Gamma_i'(J)\}}, \forall i\right\}$ is subject to the Bernoulli distribution:

$$\Pr\left(\sum_{i=1}^{K} I_{\{l\in\Gamma_i'(J)\}} \le K_1\right) = \sum_{t=0}^{K_1}\binom{K}{t}(1-p)^t p^{K-t}, \quad (60)$$

$$\Pr\left(\sum_{i=1}^{K}\max_{j\in[1,2,\cdots,M]\setminus\Gamma_c} I_{\{j\in\Gamma_i'(J)\}} \ge K_1\right) = \sum_{t=0}^{K-K_1}\binom{K}{t}(1-p)^t p^{K-t}. \quad (61)$$

Therefore, combining Eqs. (41), (42), (60), and (61), the proof of Theorem 1 is completed.

Theorem 2 provides the probability bound for common support recovery with respect to parameters K, T, P, and RICs δ_{s_c} and δ_{s_p}. From Theorem 2, we know that the proposed SJMP algorithm has a high probability of accurately identifying common support of the channel matrix, so that a better channel estimation performance can be achieved. Theorem 2 inspires us to design an algorithm with a higher common support recovery probability to achieve better channel estimation performance. The simulation in Section 5 also verifies Theorem 2.

4.2 Channel estimation quality

In this subsection, we show the channel estimation quality using our proposed SJMP algorithm for a multi-user massive MIMO system. For statistical channel model S=$\{s_c, s_p$: for K users$\}$ in Definition 1, we can obtain the following theorem:

Theorem 3 (Upper bound of the channel estimation error) In the SJMP algorithm, the estimated channel matrix for the ith user is

$$(\bar{\boldsymbol{H}}_i^{\mathrm{e}})^{[\{1,2,\cdots,M\}-\Gamma_i^{\mathrm{e}}]\to} = \boldsymbol{0}, \quad (62)$$

$$(\bar{\boldsymbol{H}}_i^{\mathrm{e}})^{\Gamma_i^{\mathrm{e}}\to} = \left(\bar{\boldsymbol{X}}_{\downarrow\Gamma_i^{\mathrm{e}}}\right)^{\dagger}\bar{\boldsymbol{Y}}_i. \quad (63)$$

Then, for the ith user, the $\boldsymbol{H}_i^{\mathrm{e}}$ obtained satisfies

$$\left\|\boldsymbol{H}_i-\boldsymbol{H}_i^{\mathrm{e}}\right\|_{\mathrm{F}} \le \frac{1}{1-\delta_{s_p}}\left\|(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus\Gamma_i^{\mathrm{e}}\to}\right\|_{\mathrm{F}} + \frac{\sqrt{1+\delta_{s_p}}}{1-\delta_{s_p}}\sqrt{\frac{TN}{P}}, \quad (64)$$

where $(\bar{\boldsymbol{H}}_i)^{\Gamma_i\setminus\Gamma_i^{\mathrm{e}}\to}$ are the residual signals based on the estimated support set Γ_i^{e} obtained for the ith user, and δ_{s_p} is the s_pth RIC of pilot sequences of $\bar{\boldsymbol{X}}$.

Proof From the transform in Eq. (23), we have

$$\left\|\boldsymbol{H}_i-\boldsymbol{H}_i^{\mathrm{e}}\right\|_{\mathrm{F}} = \left\|\bar{\boldsymbol{H}}_i-\bar{\boldsymbol{H}}_i^{\mathrm{e}}\right\|_{\mathrm{F}}. \quad (65)$$

Then we can obtain the following inequality:

$$\begin{aligned}\left\|\bar{\boldsymbol{H}}_i-\bar{\boldsymbol{H}}_i^{\mathrm{e}}\right\|_{\mathrm{F}} \leq & \left\|(\bar{\boldsymbol{H}}_i)^{\Gamma_i \backslash \Gamma_i^{\mathrm{e}} \rightarrow}\right\|_{\mathrm{F}} \\ & +\left\|(\bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}})^{\dagger} \bar{\boldsymbol{X}}_{\downarrow \Gamma_i \backslash \Gamma_i^{\mathrm{e}}}(\bar{\boldsymbol{H}}_i)^{\Gamma_i \backslash \Gamma_i^{\mathrm{e}} \rightarrow}\right\|_{\mathrm{F}}+\left\|(\bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}})^{\dagger} \bar{\boldsymbol{N}}_i\right\|_{\mathrm{F}} .\end{aligned} \tag{66}$$

Furthermore, the second term in the right-hand side of the inequality can be written as

$$\begin{aligned}& \left\|(\bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}})^{\dagger} \bar{\boldsymbol{X}}_{\downarrow \Gamma_i \backslash \Gamma_i^{\mathrm{e}}}(\bar{\boldsymbol{H}}_i)^{\Gamma_i \backslash \Gamma_i^{\mathrm{e}} \rightarrow}\right\|_{\mathrm{F}} \\ & =\left\|(\bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}}^{\mathrm{H}} \bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}})^{-1} \bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}} \bar{\boldsymbol{X}}_{\downarrow \Gamma_i \backslash \Gamma_i^{\mathrm{e}}}(\bar{\boldsymbol{H}}_i)^{\Gamma_i \backslash \Gamma_i^{\mathrm{e}} \rightarrow}\right\|_{\mathrm{F}} \\ & \leq \frac{1}{1-\delta_{s_{\mathrm{p}}}}\left\|\bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}} \bar{\boldsymbol{X}}_{\downarrow \Gamma_i \backslash \Gamma_i^{\mathrm{e}}}(\bar{\boldsymbol{H}}_i)^{\Gamma_i \backslash \Gamma_i^{\mathrm{e}} \rightarrow}\right\|_{\mathrm{F}} \\ & \leq \frac{\delta_{s_{\mathrm{p}}}}{1-\delta_{s_{\mathrm{p}}}}\left\|(\bar{\boldsymbol{H}}_i)^{\Gamma_i \backslash \Gamma_i^{\mathrm{e}} \rightarrow}\right\|_{\mathrm{F}} .\end{aligned} \tag{67}$$

The third term is

$$\begin{aligned}& \left\|(\bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}})^{\dagger} \bar{\boldsymbol{N}}_i\right\|_{\mathrm{F}}=\left\|(\bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}}^{\mathrm{H}} \bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}})^{-1} \bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}} \bar{\boldsymbol{N}}_i\right\|_{\mathrm{F}} \\ & \leq \frac{1}{1-\delta_{s_{\mathrm{p}}}}\left\|\bar{\boldsymbol{X}}_{\downarrow \Gamma_i^{\mathrm{e}}} \bar{\boldsymbol{N}}_i\right\|_{\mathrm{F}} \leq \frac{\sqrt{1+\delta_{s_{\mathrm{p}}}}}{1-\delta_{s_{\mathrm{p}}}}\left\|\bar{\boldsymbol{N}}_i\right\|_{\mathrm{F}} \\ & =\frac{\sqrt{1+\delta_{s_{\mathrm{p}}}}}{1-\delta_{s_{\mathrm{p}}}} \sqrt{\frac{TN}{P}} .\end{aligned} \tag{68}$$

Substituting Eqs. (66)–(68) into Eq. (65), Theorem 3 is proved.

Theorem 3 provides the channel estimation error upper bound of each user using our proposed SJMP algorithm with respect to parameters N, T, P, and RIC $\delta_{s_{\mathrm{p}}}$. Note that the term $(\bar{\boldsymbol{H}}_i)^{\Gamma_i \backslash \Gamma_i^{\mathrm{e}} \rightarrow}$ represents the residual signals based on the estimated support set Γ_i^{e} obtained by the SJMP algorithm for the ith user. The concrete meaning is the rth row vector of $\bar{\boldsymbol{H}}_i$ in support set $\Gamma_i \backslash \Gamma_i^{\mathrm{e}}$, where the index r satisfies $r \in \Gamma_i, r \notin \Gamma_i^{\mathrm{e}}$. To be specific, if $\Gamma_i=\Gamma_i^{\mathrm{e}}$, the support set is estimated accurately and there is no element in support set $\Gamma_i \backslash \Gamma_i^{\mathrm{e}}$, so that $(\bar{\boldsymbol{H}}_i)^{\Gamma_i \backslash \Gamma_i^{\mathrm{e}} \rightarrow}=\mathbf{0}$. Although the Γ_i^{e} cannot be exactly correct compared with Γ_i, according to Theorem 1, the correct common support Γ_{c} can be obtained with a high probability $\Pr(E_{\mathrm{C}})$, leading to a high probability of accurately recovering Γ_i^{e}. As a result, we can use Eq. (64) as the upper bound of the channel estimation error.

4.3 Computational complexity

In this subsection, the computational complexity of the proposed SJMP is analyzed. The main complexity in each iteration comes from several operations as shown below. For step 2, the correlation and l_2-norm in the support estimate have a computational complexity of $O(TMN)$ and $O(MN)$, respectively. For step 3, the function $\arg\max(\cdot)$ has a complexity of $O(Ks_{\mathrm{p}})$. For step 4, the overall complexity in the residue update is $O(s_{\mathrm{c}}^3+2Ts_{\mathrm{c}}^2+T^2(s_{\mathrm{c}}+N))$, including pseudo inversion $O(s_{\mathrm{c}}^3+2Ts_{\mathrm{c}}^2)$ and residue update $O(T^2(s_{\mathrm{c}}+N))$. For step 6, the complexity of the support update is similar to that in step 2, i.e., $O(TMN)$. In step 7, the overall computational complexity is $O(8s_{\mathrm{p}}^3+8Ts_{\mathrm{p}}^2+2NTs_{\mathrm{p}})$, including the complexities of the pseudo inversion and matrix product. For step 8, the norm operation contains a complexity of $O(2Ns_{\mathrm{p}})$. Also, the residual update in step 9 has a complexity of $O(s_{\mathrm{p}}^3+2Ts_{\mathrm{p}}^2+T^2(s_{\mathrm{p}}+N))$, which is similar to that in step 4. In step 10, the main complexity is in pseudo inversion and matrix product, with a complexity of $O(s_{\mathrm{p}}^3+2Ts_{\mathrm{p}}^2+TNs_{\mathrm{p}})$. Therefore, the overall complexity of the proposed SJMP algorithm is $O(s_{\mathrm{p}}^3+2Ts_{\mathrm{p}}^2+T^2(s_{\mathrm{p}}+N)+TMN)$. For clarity, the computational complexities of the proposed SJMP algorithm are summarized in Table 1.

Table 1 Computational complexities of the SJMP algorithm

Implementation step	Complexity
Step 2	$O(TMN)$
Step 3	$O(Ks_{\mathrm{p}})$
Step 4	$O(s_{\mathrm{c}}^3+2Ts_{\mathrm{c}}^2+T^2(s_{\mathrm{c}}+N))$
Step 6	$O(TMN)$
Step 7	$O(8s_{\mathrm{p}}^3+8Ts_{\mathrm{p}}^2+2NTs_{\mathrm{p}})$
Step 8	$O(2Ns_{\mathrm{p}})$
Step 9	$O(s_{\mathrm{p}}^3+2Ts_{\mathrm{p}}^2+T^2(s_{\mathrm{p}}+N))$
Step 10	$O(s_{\mathrm{p}}^3+2Ts_{\mathrm{p}}^2+TNs_{\mathrm{p}})$
Overall	$O(s_{\mathrm{p}}^3+2Ts_{\mathrm{p}}^2+T^2(s_{\mathrm{p}}+N)+TMN)$

5 Simulation results

The channel estimation performance of the proposed SJMP algorithm for a multi-user massive MIMO system was verified by a series of experiments. The conventional LS estimation, OMP algorithm, SOMP algorithm, and J-OMP algorithm were considered as benchmarks. The practical channel sparsity $\Gamma=\{\Gamma_c,\ \{\Gamma_i:\ \forall i\},\ 1\le i\le K\}$ was randomly generated. Concretely, the sparsity levels $|\Gamma_c|$ and $|\Gamma_i|$ were generated uniformly in $[s_c,\ s_c+2]$ and $[s_p-2,\ s_p]$, respectively. Channel coefficients of spatial paths were generated using the SCM channel model (Baum *et al.*, 2005). The pilot overhead ratio is defined as $\eta_p=T/M$. The average mean squared error (MSE) of the estimated channel is given by

$$\mathrm{MSE}=\frac{1}{K}\sum_{i=1}^{K}\frac{\left\|\boldsymbol{H}_i-\boldsymbol{H}_i^{e}\right\|_2^2}{\left\|\boldsymbol{H}_i\right\|_2^2}. \tag{69}$$

Fig. 3 presents the MSE performance of channel estimation using the proposed SJMP algorithm under the condition that users are equipped with 1, 2, or 4 antennas. Parameters are set as follows: the number of BS antennas M=200, user number K=30, common sparsity s_c=6, individual sparsity s_p=12, and pilot overhead T=45. The MSE of channel estimation decreased with the SNR. Moreover, as the number of user antennas doubled, the MSE performance gained 2.5 dB in terms of SNR. This is because users were equipped with more than one antenna, the support of the channel matrix among different user antennas was identical, leading to either all zero or all nonzero of the channel matrix. Furthermore, the block coherence of the matrix in Eq. (20) was smaller than the mutual coherence of the pilot sequence matrix. Hence, a better estimation performance was achieved, which validates the result in Theorem 1. In the following experiments, each user was equipped with two antennas.

Fig. 4 compares the MSE performance of channel estimation versus overhead T among LS estimation, OMP algorithm, SOMP algorithm, J-OMP algorithm, and our proposed SJMP algorithm, with K=30, M=200, s_c=6, s_p=12, and SNR P=20 dB. The LS algorithm performed poorly. The proposed SJMP algorithm outperformed the other four algorithms because the common sparsity of the statistical channel is leveraged. Moreover, as the overhead T increased, the MSE performance improved. The reason is that the recovery probability of the CS-based algorithm is higher with more overhead. Furthermore, when compared with the J-OMP algorithm, the performance of the proposed SJMP was consistently superior since the support estimate in step 2 provides more flexible atom selection, which results in a higher common support recovery probability. Although only η_p=20% pilot overhead was used, the channel estimation performance was still significantly superior to that of conventional LS estimation, OMP, SOMP, and J-OMP algorithms.

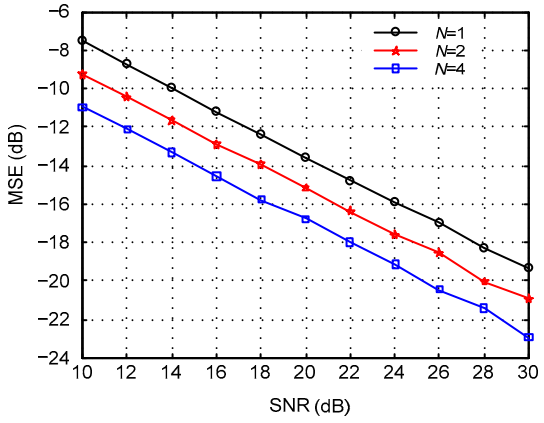

Fig. 3 MSE performance with different numbers of user antennas

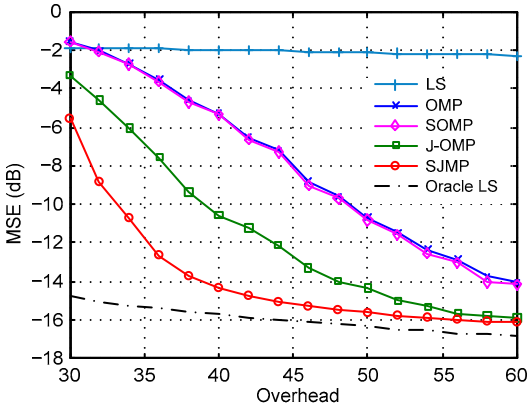

Fig. 4 MSE performance of channel estimation at different overhead *T*

Fig. 5 illustrates the MSE performance with common sparsity s_c corresponding to parameters K=30, M=200, s_p=12, SNR P=20 dB, and T=45. As the common sparsity increased, the MSE using the

other four algorithms remained invariant, while that of our proposed SJMP decreased slightly because steps 2–4 tend to identify the common support correctly.

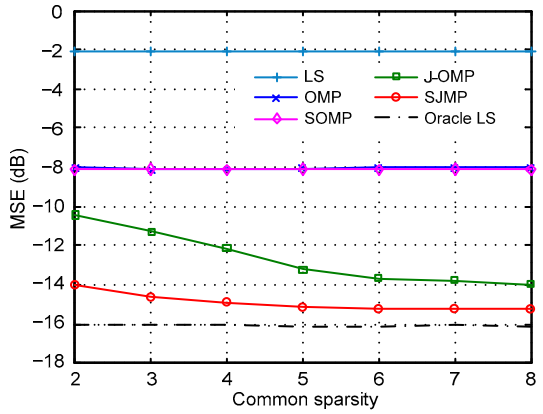

Fig. 5 MSE performance of channel estimation at different common sparsity s_c

Fig. 6 depicts the MSE performance with individual sparsity s_p corresponding to parameters K=30, M=200, s_c=6, SNR P=20 dB, and T=45. The MSE increased with the individual sparsity. The reason is that high sparsity requires more measurements, which is equivalent to requiring more overhead T. Thus, given a certain pilot overhead, the channel estimation MSE becomes worse as the individual sparsity increases.

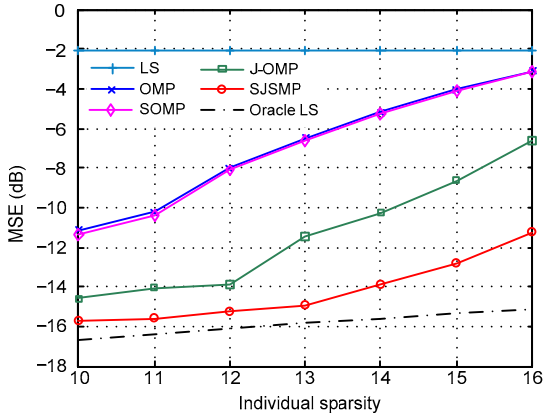

Fig. 6 MSE performance of channel estimation at different individual sparsity s_p

Fig. 7 shows the MSE performance versus the number of BS antennas M with parameters K=30, s_c=6, s_p=12, T=45, and P=20 dB. As the number of BS antennas M grows, the MSE of the proposed algorithm rises slightly. As in Fig. 4, more antennas require more overhead T. When the pilot overhead remained constant, the MSE performance became worse with more BS antennas. Note that the proposed SJMP algorithm outperformed the other algorithms. The reason is that the common support and individual support of the channel matrix can be identified accurately.

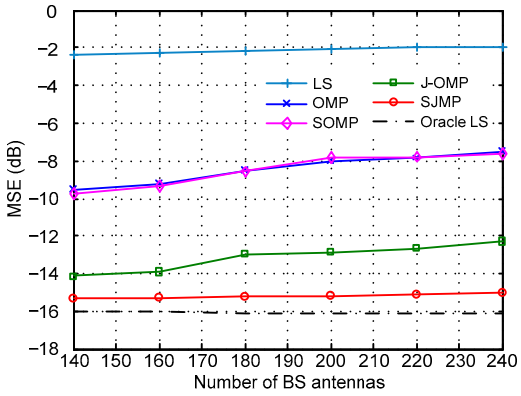

Fig. 7 MSE performance of channel estimation at different BS antenna numbers

Fig. 8 shows the MSE performance versus SNR P with K=30, M=200, s_c=6, s_p=12, and T=45. The channel estimation performance of all the algorithms improved as SNR increased. More importantly, the proposed SJMP consistently performed better than the other algorithms with increasing SNR.

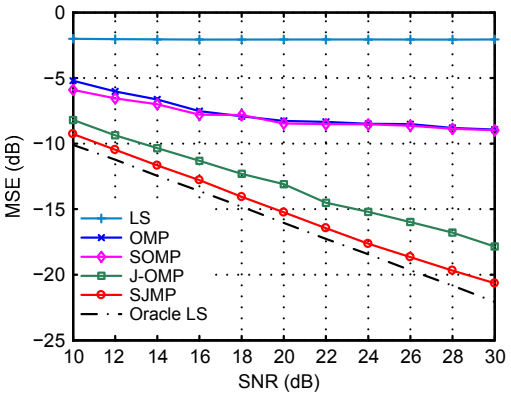

Fig. 8 MSE performance of channel estimation at different SNR

Fig. 9 compares the MSE performance wiht the number of users K with parameters M=200, s_c=6, s_p=12, SNR P=20 dB, and T=45. The channel esti-

mation performance improved with the increasing number of users. Note from Eq. (12) that a larger K provides more common sparsity of the user channel matrix, so that the proposed SJMP can exploit the structure of the channel sparsity of all users, achieving better estimation performance. Note that a cross curve occurred between J-OMP, OMP, and SOMP. Furthermore, when the number of users was small, the MSE performance of the J-OMP algorithm was even worse than that of OMP and SOMP. This is because J-OMP cannot obtain the common support among different user channel matrices with fewer users. As the number of users increased, the performance of J-OMP improved. This is mainly because J-OMP can exploit the common support of the channel matrix, while the OMP and SOMP algorithms fail to take advantage of that property.

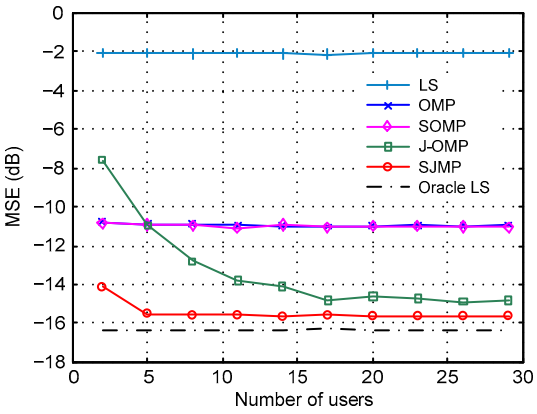

Fig. 9 MSE performance of channel estimation at different user numbers

Note that from Figs. 4–9, the performance of the proposed SJMP was significantly better than that of OMP, SOMP, and J-OMP. The reason is that the proposed SJMP achieves a higher probability of sparsity support recovery. Fig. 10 illustrates the success rate of common support recovery versus the pilot overhead. The simulation parameters were set as K=30, M=200, s_c=6, s_p=12, and P=20 dB. Each test point was run 100 times and successful common support recovery was declared when the true common support was the subset of the estimated common support. The success rate increased with overhead T (Fig. 10), following the same trend as in Fig. 4, leading to better channel estimation of each user. In particular, for the proposed SJMP algorithm, using only η_p=22% pilot overhead, the success rate was about 90%, which guarantees accurate channel estimation for the multi-user massive MIMO system. When using η_p=28% pilot overhead, the success rate was nearly 100%, which validates the results in Fig. 4. Furthermore, the results of Figs. 4 and 9 verify Theorem 2, where a high probability of recovering common support leads to better channel estimation performance.

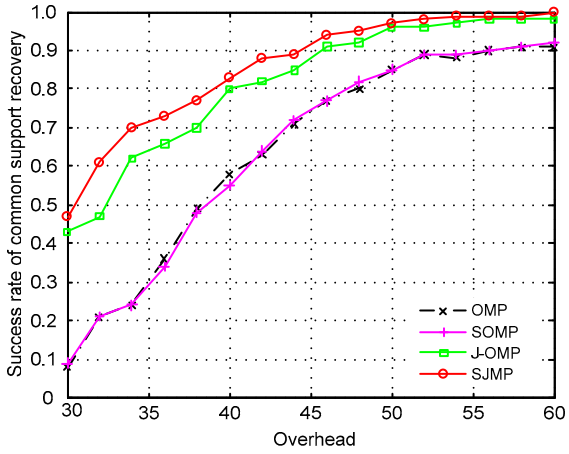

Fig. 10 Success rate of common support recovery versus overhead T

Besides comparing the MSE performance of the proposed SJMP algorithm under various simulation settings, the bit error rate (BER) performance and average achievable rate per user performance were investigated. Fig. 11 depicts the downlink BER performance versus SNR P, with the parameters set as K=30, M=200, s_c=6, s_p=12, and T=50. The estimated channels of four benchmark algorithms and the proposed SJMP algorithms were used by the BS to perform zero-forcing (ZF) precoding, so that the BS could simultaneously serve K users using 16-quadrature amplitude modulation as the transmission signal. The proposed SJMP algorithm outperformed the other four algorithms because the common sparsity of the statistical channel is leveraged, which can be verified by Figs. 5 and 8. When compared with the J-OMP algorithm, the performance gain of the proposed SJMP was about 0.8 dB in terms of BER with only 0.2 dB performance loss compared with the Oracle LS algorithm. We conclude that the better MSE performance of channel estimation leads to more effective information transmission.

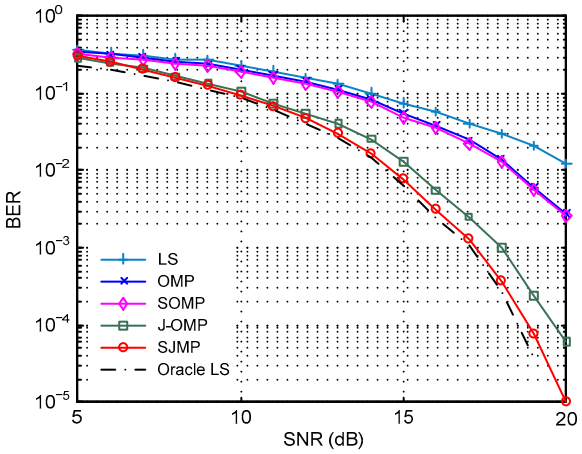

Fig. 11 BER performance at different SNR *P*

In Fig. 12, when the estimated channels of different algorithms were employed, the average achievable rate per user performance was investigated with the same parameters in Fig. 11. At the BS, the precoding vector of user i for the different algorithms was generated according to the corresponding estimated channels, i.e., $\boldsymbol{W}_i^{\mathrm{e}}=\left(\boldsymbol{H}_i^{\mathrm{e}}\right)^{\dagger}\Big/\left\|\left(\boldsymbol{H}_i^{\mathrm{e}}\right)^{\dagger}\right\|$. Then the signal-to-interference-plus-noise ratio (SINR) of the ith user can be denoted as

$$\mathrm{SINR}_i=\frac{P}{K}\left|\boldsymbol{H}_i\boldsymbol{W}_i^{\mathrm{e}}\right|^2\Big/\left(1+\frac{P}{K}\sum_{m=1,m\neq i}^{K}\left|\boldsymbol{H}_i\boldsymbol{W}_m^{\mathrm{e}}\right|^2\right). \quad (70)$$

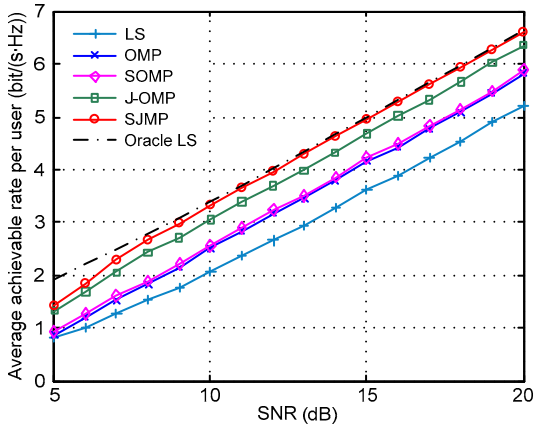

Fig. 12 Average achievable rate per user performance at different SNR *P*

The average achievable rate per user can be calculated as $E[\log_2(1+\mathrm{SINR}_i)]$, where $E[\cdot]$ represents the mean value for the achievable rate of all the K users. The achievable rate increased with SNR (Fig. 12). In particular, the proposed SJMP consistently performed better than the other algorithms because of its more accurate channel estimation. Moreover, since the individual support of the channel matrix of each user can be precisely acquired, the gap between the proposed SJMP algorithm and perfect channel information for the average achievable rate can be negligible.

6 Conclusions

In this paper, to reduce the pilot overhead in massive MIMO systems, we have proposed a structured joint channel estimation scheme deploying CS theory. In particular, due to the practical scattering environment, the channel sparsity in the angular domain was analyzed. The common sparsity and individual sparsity structure among neighboring users exist in the channel matrix. Then, users equipped with multiple antennas were investigated in the framework of CS to further reduce the pilot overhead. Consequently, by leveraging the statistical sparsity structure, the SJMP algorithm at the BS was proposed to jointly estimate channels associated with limited pilot overhead. Also, the probability upper bound of the common sparsity recovery and the upper bound of the channel estimation quality of the proposed SJMP algorithm were deduced, providing a vision to design algorithms to further reduce pilot overhead and improve channel performance. Simulations showed that the proposed algorithm achieves superior channel estimation performance with a lower pilot overhead.

References

Barbotin, Y., Hormati, A., Rangan, S., *et al.*, 2012. Estimation of sparse MIMO channels with common support. *IEEE Trans. Commun.*, **60**(12):3705-3716. https://doi.org/10.1109/TCOMM.2012.091112.110439

Baum, D.S., Hansen, J., Salo, J., 2005. An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM). IEEE 61st Vehicular Technology Conf., p.3132-3136. https://doi.org/10.1109/VETECS.2005.1543924

Berger, C.R., Wang, Z.H., Huang, J.Z., *et al.*, 2010. Application of compressive sensing to sparse channel estimation. *IEEE Commun. Mag.*, **48**(11):164-174. https://doi.org/10.1109/MCOM.2010.5621984

Björnson, E., Larsson, E.G., Marzetta, T.L., 2015. Massive MIMO: ten myths and one critical question. *IEEE Commun. Mag.*, **54**(2):114-123.

https://doi.org/10.1109/MCOM.2016.7402270

Bogale, T.E., Vandendorpe, L., Chalise, B.K., 2012. Robust transceiver optimization for downlink coordinated base station systems: distributed algorithm. *IEEE Trans. Signal Process*., **60**(1):337-350. https://doi.org/10.1109/TSP.2011.2170167

Chen, Y., Qin, Z., 2015. Gradient-based compressive image fusion. *Front. Inform. Technol. Electron. Eng.*, **16**(3):227-237. https://doi.org/10.1631/FITEE.1400217

Choi, J., Love, D.J., Bidigare, P., 2014. Downlink training techniques for FDD massive MIMO systems: open-loop and closed-loop training with memory. *IEEE J. Sel. Top. Signal Process.*, **8**(5):802-814. https://doi.org/10.1109/JSTSP.2014.2313020

Dai, L.L., Wang, J.T., Wang, Z.C., *et al*., 2013. Spectrum- and energy-efficient OFDM based on simultaneous multi-channel reconstruction. *IEEE Trans. Signal Process.*, **61**(23):6047-6059. https://doi.org/10.1109/TSP.2013.2282920

Dai, W., Milenkovic, O., 2009. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inform. Theory*, **55**(5):2230-2249. https://doi.org/10.1109/TIT.2009.2016006

Dasgupta, S., Gupta, A., 2003. An elementary proof of a theorem of Johnson and Lindenstrauss. *Rand. Struct. Algor.*, **22**(1):60-65. https://doi.org/10.1002/rsa.10073

Donoho, D.L., 2006. Compressed sensing. *IEEE Trans. Inform. Theory*, **52**(4):1289-1306. https://doi.org/10.1109/TIT.2006.871582

Eldar, Y.C., Kuppinger, P., Bölcskei, H., 2010. Block-sparse signals: uncertainty relations and efficient recovery. *IEEE Trans. Signal Process.*, **58**(6):3042-3054. https://doi.org/10.1109/TSP.2010.2044837

Gao, X., Edfors, O., Rusek, F., *et al*., 2011. Linear pre-coding performance in measured very-large MIMO channels. IEEE Vehicular Technology Conf., p.1-5. https://doi.org/10.1109/VETECF.2011.6093291

Gao, Z., Dai, L.L., Wang, Z., 2014. Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems. *Electron. Lett.*, **50**(12):896-898. https://doi.org/10.1049/el.2014.0985

Gao, Z., Dai, L.L., Wang, Z., *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., Dai, L.L., Dai, W., *et al*., 2016. Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO. *IEEE Trans. Commun.*, **64**(2):601-617. https://doi.org/10.1109/TCOMM.2015.2508809

Hoydis, J., Hoek, C., Wild, T., *et al*., 2012. Channel measurements for large antenna arrays. IEEE Int. Symp. on Wireless Communication Systems, p.811-815. https://doi.org/10.1109/ISWCS.2012.6328480

Hoydis, J., Ten Brink, S., Debbah, M., 2013. Massive MIMO in the UL/DL of cellular networks: how many antennas do we need? *IEEE J. Sel. Areas Commun.*, **31**(2):160-171. https://doi.org/10.1109/JSAC.2013.130205

Hu, D., Wang, X.D., He, L.H., 2013. A new sparse channel estimation and tracking method for time-varying OFDM systems. *IEEE Trans. Veh. Technol.*, **62**(9):4648-4653. https://doi.org/10.1109/TVT.2013.2266282

Ketonen, J., Juntti, M., Cavallaro, J.R., 2010. Performance-complexity comparison of receivers for a LTE MIMO-OFDM system. *IEEE Trans. Signal Process.*, **58**(6):3360-3372. https://doi.org/10.1109/TSP.2010.2044290

Lee, B., Choi, J., Seol, J.Y., *et al*., 2015. Antenna grouping based feedback compression for FDD-based massive MIMO systems. *IEEE Trans. Commun.*, **63**(9):3261-3274. https://doi.org/10.1109/TCOMM.2015.2460743

Lu, L., Li, G.Y., Swindlehurst, A.L., *et al*., 2014. An overview of massive MIMO: benefits and challenges. *IEEE J. Sel. Top. Signal Process.*, **8**(5):742-758. https://doi.org/10.1109/JSTSP.2014.2317671

Noh, S., Zoltowski, M.D., Sung, Y., *et al*., 2014. Pilot beam pattern design for channel estimation in massive MIMO systems. *IEEE J. Sel. Top. Signal Process.*, **8**(5):787-801. https://doi.org/10.1109/JSTSP.2014.2327572

Qi, C.H., Wu, L.N., 2014. Uplink channel estimation for massive MIMO systems exploring joint channel sparsity. *Electron. Lett.*, **50**(23):1770-1772. https://doi.org/10.1049/iet-com.2013.0781

Rao, X.B., Lau, V.K.N., 2014. Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. *IEEE Trans. Signal Process*., **62**(12):3261-3271. https://doi.org/10.1109/TSP.2014.2324991

Tropp, J.A., Gilbert, A.C., 2007. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, **53**(12):4655-4666. https://doi.org/10.1109/TIT.2007.909108

Tropp, J.A., Gilbert, A.C., Strauss, M.J., 2006. Algorithms for simultaneous sparse approximation. Part I: greedy pursuit. *Signal Process.*, **86**(3):572-588. https://doi.org/10.1016/j.sigpro.2005.05.030

Tse, D., Viswanath, P., 2005. Fundamentals of Wireless Communication. Cambridge University Press, New York, p.309-330.

Yin, H.F., Gesbert, D., Filippou, M., *et al*., 2012. A coordinated approach to channel estimation in large-scale multiple-antenna systems. *IEEE J. Sel. Areas Commun.*, **31**(2):264-273. https://doi.org/10.1109/JSAC.2013.130214

Zhang, Z.Y., Teh, K.C., Li, K.H., 2014. Application of compressive sensing to limited feedback strategy in large-scale multiple-input single-output cellular networks. *IET Commun.*, **8**(6):947-955. https://doi.org/10.1049/iet-com.2013.0781

Appendix: Proof of Theorem 1

Given that the pilot sequence matrix is normalized, Eq. (19) can be simplified as

$$\mu(\boldsymbol{X}'^{\mathrm{T}})=\max_{\substack{1\le i,l\le M\\ i\ne l}}\left|\left\langle \boldsymbol{x}'_i,\boldsymbol{x}'_l\right\rangle\right|. \tag{A1}$$

To explain it clearly, the concrete form of pilot matrix $\boldsymbol{X}'^{\mathrm{T}}$ is expressed as

$$\boldsymbol{X}'^{T}=\begin{bmatrix} x_{11} & x_{1i} & \cdots & x_{1l} & x_{1M}\\ \vdots & \vdots & & \vdots & \vdots\\ x_{T1} & \underbrace{x_{Ti}}_{\boldsymbol{x}'_i} & \cdots & \underbrace{x_{Tl}}_{\boldsymbol{x}'_l} & x_{TM}\end{bmatrix}. \tag{A2}$$

According to Eq. (18), $\boldsymbol{X}'_{\mathrm{B}}$ is expressed as Eq. (A3).

It is clear that the first column of $\boldsymbol{x}'_{\mathrm{B}i}$ consists of $\boldsymbol{x}'_i$ with interpolating $N-1$ zeros between two adjacent elements. The rest of column of $\boldsymbol{x}'_{\mathrm{B}i}$ is constituted by a cyclic shift of the first column. $\boldsymbol{x}'_{\mathrm{B}l}$ is made up in a similar way.

According to the definition of 'block' coherence in Eq. (20), we have

$$\begin{aligned}\mu_{\mathrm{B}}(\boldsymbol{X}'_{\mathrm{B}})&=\frac{1}{N}\max_{\substack{1\le i,l\le M\\ i\ne l}}\rho\left(\boldsymbol{x}'^{\mathrm{H}}_{\mathrm{B}i}\boldsymbol{x}'_{\mathrm{B}l}\right)\\ &=\frac{1}{N}\max_{\substack{1\le i,l\le M\\ i\ne l}}\rho\left(\mathrm{diag}\left\{\underbrace{(\boldsymbol{x}'_i)^{\mathrm{H}}\boldsymbol{x}'_l,(\boldsymbol{x}'_i)^{\mathrm{H}}\boldsymbol{x}'_l,\cdots,(\boldsymbol{x}'_i)^{\mathrm{H}}\boldsymbol{x}'_l}_{N}\right\}\right).\end{aligned} \tag{A4}$$

Applying Eq. (21), we have Eq. (A5).

$$\boldsymbol{X}'_{\mathrm{B}}=\begin{bmatrix} x_{11}\boldsymbol{I}_N & x_{1i}\boldsymbol{I}_N & \cdots & x_{1l}\boldsymbol{I}_N & x_{1M}\boldsymbol{I}_N\\ \vdots & \vdots & & \vdots & \vdots\\ x_{T1}\boldsymbol{I}_N & \underbrace{x_{Ti}\boldsymbol{I}_N}_{\boldsymbol{x}'_{\mathrm{B}i}} & \cdots & \underbrace{x_{Tl}\boldsymbol{I}_N}_{\boldsymbol{x}'_{\mathrm{B}l}} & x_{TM}\boldsymbol{I}_N\end{bmatrix}=\begin{bmatrix} & \overbrace{\begin{matrix}x_{1i}&0&0\end{matrix}}^{N\text{ columns}} & & \overbrace{\begin{matrix}x_{1l}&0&0\end{matrix}}^{N\text{ columns}} & \\ x_{11}\boldsymbol{I}_N & \begin{matrix}0&&0\end{matrix} & \cdots & \begin{matrix}0&&0\end{matrix} & x_{1M}\boldsymbol{I}_N\\ & \begin{matrix}0&0&x_{1i}\end{matrix} & & \begin{matrix}0&0&x_{1l}\end{matrix} & \\ \vdots & \vdots & & \vdots & \vdots\\ & \begin{matrix}x_{Ti}&0&0\end{matrix} & & \begin{matrix}x_{Tl}&0&0\end{matrix} & \\ x_{T1}\boldsymbol{I}_N & \begin{matrix}0&&0\end{matrix} & \cdots & \begin{matrix}0&&0\end{matrix} & x_{TM}\boldsymbol{I}_N\\ & \underbrace{\begin{matrix}0&0&x_{Ti}\end{matrix}}_{\boldsymbol{x}'_{\mathrm{B}i}} & & \underbrace{\begin{matrix}0&0&x_{Tl}\end{matrix}}_{\boldsymbol{x}'_{\mathrm{B}l}} & \end{bmatrix}. \tag{A3}$$

$$\mu_{\mathrm{B}}(\boldsymbol{X}'_{\mathrm{B}})=\frac{1}{N}\max_{\substack{1\le i,l\le M\\ i\ne l}}\max_{k}\sqrt{\lambda_k\left(\mathrm{diag}\left(\underbrace{\left|(\boldsymbol{x}'_i)^{\mathrm{H}}\boldsymbol{x}'_l\right|^2,\left|(\boldsymbol{x}'_i)^{\mathrm{H}}\boldsymbol{x}'_l\right|^2,\cdots,\left|(\boldsymbol{x}'_i)^{\mathrm{H}}\boldsymbol{x}'_l\right|^2}_{N}\right)\right)}=\frac{1}{N}\mu(\boldsymbol{X}'^{\mathrm{T}}). \tag{A5}$$