# Threshold-Enhanced Hierarchical Spatial Non-Stationary Channel Estimation for Uplink Massive MIMO Systems

Chongyang Tan, *Student Member, IEEE*, Donghong Cai, *Member, IEEE*, Yanqing Xu, *Member, IEEE*,
Zhiguo Ding, *Fellow, IEEE*, and Pingzhi Fan, *Fellow, IEEE*

*Abstract*—Spatial non-stationarity channel estimation for u-plink massive MIMO systems can be formulated as a non-uniform block sparse signal recovery problem, in which the hierarchical sparsity of the channel matrix is classified as row sparsity and in-row sparsity. This paper proposes an efficient threshold-enhanced hierarchical estimation (TEHE) algorithm without prior information. More precisely, the non-zero rows of spatial non-stationarity channel matrix are estimated according to the in-row correlation in the first layer; while the non-zero elements of the estimated non-zero rows are further refined in the second layer. Different from the existing two-layer iteration algorithms, an adaptive threshold is designed to estimate the non-zero elements replacing the iterative algorithm in the second layer. In the proposed TEHE algorithm, row-wise sparse adaptive matching pursuit (SAMP) is used to find the non-zero rows in the first layer, which has high precision and lower complexity, compared to the conventional SAMP. To further improve the efficiency of the row estimation for larger antenna array, an adaptive threshold-enhanced hierarchical estimation (A-TEHE) algorithm is proposed. In addition, a sufficient condition and a halting condition for theoretical guarantee to obtain accurate row estimation are developed. Finally, the computation complexity is analyzed and compared. The simulation results demonstrate that the proposed threshold-enhanced hierarchical spatial non-stationary channel estimation algorithms achieve better performance compared to various state-of-the-art baselines in terms of support set estimation, channel coefficient estimation, and computational efficiency. Specifically, the proposed algorithms are robust to the in-row sparsity.

*Index Terms*—patialpatialS non-stationary channel estimation, compressive sensing, hierarchical estimation, massive MIMO.

C. Tan and D. Cai are with the College of Information Science Technology, Jinan University, Guangzhou 510632, China (e-mail: cyatan@stu2021.jnu.edu.cn; dhcai@jnu.edu.cn).

Y. Xu is with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen 518172, China (e-mail: xuyanqing@cuhk.edu.cn).

Z. Ding is with the School of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, U.K. (e-mail: zhiguo.ding@manchester.ac.uk).

P. Fan is with n is with the Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (e-mail: pzfan@swjtu.edu.cn).

## I. INTRODUCTION

WITH the rapid development of the Internet of Things (IoT) [2] in the fifth generation (5G) networks [3]–[5], such as massive machine-type communication, autonomous driving, future communication networks have higher requirements of spectral efficiency, energy efficiency, and transmission/processing delay, etc.. Massive/Extremely-large-scale multiple-input multiple-output (MIMO) system increases the antenna scale of the base station (BS) to improve the throughput and expands the coverage by utilizing spatial freedom. Meanwhile, a large number of BS antennas bring high array gain, which can significantly reduce the power consumption of each user. Besides, a combination of massive MIMO and intelligent reflecting surfaces [6], distributed MIMO antenna array systems [7], and intelligent signal processing of extremely large-scale antenna array systems [8] can further optimize the efficiency of massive MIMO systems in 5G. Consequently, the MIMO system equipped with more BS intelligent antennas is regarded as a promising technology towards the sixth generation (6G) communication networks [9]–[11]. In practical massive MIMO systems, lower pilot cost and accurate channel information acquisition by the BS are the keys to implementing simple linear precoding [12] and approaching the upper bound of system performance. However, the pilot cost of channel estimation increases with the increase of the number of transmitting antennas or the served users, and uplink channel estimation becomes more challenging.

To reduce the pilot cost and the complexity of massive MIMO channel estimation, one of the main methods is to optimize the channel estimation algorithm based on the hidden sparsity [13]–[15]. For example, the channel matrix can be sparsely represented in the angular domain or polarization domain. In particular, the sparse patterns of massive MIMO channels can be sorted into two types: spatial stationary and non-stationary characteristics, whose channel impulse responses (CIRs) are shown in the following Fig. 1. For the spatial stationary channel, as Fig. 1(a) shown, all of the elements in each row of the channel matrix are either zeros or non-zeros, i.e., the spatial stationary channel has only all-zero and all-non-zero rows. However, in the spatial non-stationary channel, there is a different sparse structure of the BS antennas, and the sparsity of the whole channel becomes non-uniform, which is reflected in Fig. 1(b) with rows containing both zero and non-

(a) CIRs of spatial stationary channel.

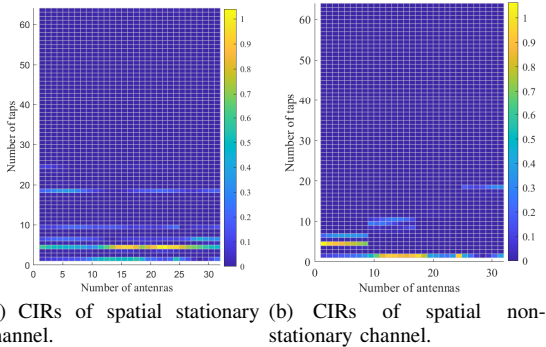(b) CIRs of spatial non-stationary channel.

Fig. 1. Illustration of CIRs for spatial stationary and non-stationary channels.

zero elements. It is more challenging to design spatial non-stationary channel estimation algorithms due to the special non-uniform block sparse structure. Most recently, the works on spatial non-stationary channel estimation have attracted more attention [16]–[18]. In [16], a sparse Bayesian algorithm based on hierarchical Dirichlet process prior information is proposed for uplink spatial non-stationary channel estimation. Further, a three-layer Markov chain is adopted to formulate the visibility and delay prior information in [19], and a turbo orthogonal approximate message passing algorithm is adopted to estimate the uplink channel of MIMO with an extremely large-scale antenna array. In addition, a pattern-coupled prior is used to formulate the non-uniform sparse structure of massive MIMO channels [20], and a generic sparse Bayesian learning framework is developed. These works make reasonable use of channel sparse structure for estimation. However, prior information is required and the complexity limits the antenna array size.

Recently, compressed sensing (CS) based techniques [21]–[26] are applied to the massive MIMO channel estimations. Unlike the Bayesian-type estimation algorithms, which rely on the prior distribution of the channel, the CS-based techniques take advantage of the special sparse structure of channels. For example, an alternative CS-based channel estimator for millimeter wave hybrid MIMO systems is developed in [27], in which grid-based orthogonal matching pursuit is proposed. Moreover, a large aperture array is divided into multiple subarrays and the channel of last-hop scatterers is modeled as a spherical wavefront in [28]. Then, the subarray-wise and the scatterer-wise channel estimation methods are proposed, respectively. Further, [29] proposes a generalized block orthogonal matching pursuit estimation framework by considering the special block sparse structure of the channel matrix. However, this characteristic only appears in the stationary channel. Thus, the block CS-based algorithm is inefficient for spatial non-stationary channel estimation. In [30], a two-stage sparse channel estimation scheme based on block matching pursuit (BMP) is proposed for spatial non-stationary channel estimation. The estimation accuracy is improved by exploiting the non-uniform block sparse structure of spatial non-stationary channels, while the complexity is high due to two iteration stages with BMP.

In this paper, we consider the uplink spatial non-stationary

channel estimation of massive MIMO systems, where the hierarchical sparse structure is used to model the channel matrix, i.e., row sparsity, and in-row sparsity. Unlike the general block sparse recovery problem, we consider an non-uniform block sparse matrix and model it as a hierarchical sparse recovery problem. Two threshold-enhanced hierarchical estimation algorithms comprised of two-layer estimations are proposed without prior information. Different from the first layer of the traditional two-stage iteration algorithm, a block sparse adaptive matching pursuit (SAMP) is used to estimate the non-zeros rows of the channel matrix in the first layer, while an adaptive threshold is designed for refining the elements of the estimated non-zero rows. More importantly, the proposed threshold for refining the elements in the non-zero rows can still work well even though the in-row sparsity is poor. In addition, the performance of the proposed algorithm, including estimation accuracy, halting condition and computation complexity are analyzed. The main contributions of this paper are summarized as follows:

- A threshold enhanced hierarchical estimation (TEHE) algorithm including two layers is proposed for uplink spatial non-stationary channel estimation of massive MIMO systems. Specifically, a block SAMP algorithm based on the in-row correlation of the channel matrix is used to estimate the non-zero rows in the first layer, and a threshold based on the reconstructed noise level is designed for element refining in the second layer.
- To further improve the efficiency of the proposed TEHE algorithm, an adaptive threshold- enhanced hierarchical estimation (A-TEHE) algorithm is proposed. Higher efficiency can be obtained by adjusting the step size of the first layer, and the normalized mean square error (NMSE) performance can be improved for low signal-to-noise ratio (SNR) region. In addition, the designed threshold can be used to find out the support set and refine the estimated channel effectively. Different from the existing two-stage iteration algorithms, the proposed algorithm can be applied to enhance the estimated non-zero channel coefficient with different sparsity patterns. Even if it is close to the dense case, i.e., nearly spatial stationary channel, the estimation performance can still be guaranteed.
- The theoretical guarantees to achieve the desired accuracy in the first layer estimation in the proposed TEHE algorithm are derived. The developed sufficient conditions prove that the proposed algorithm can correctly find all row support sets of the sparse channel matrix. Moreover, the stopping conditions and the complexity of the proposed algorithms are analyzed and discussed.

Simulation results indicate that the proposed threshold-enhanced hierarchical spatial non-stationary channel estimation algorithms outperform various state-of-the-art baselines in terms of support set estimation, channel coefficient estimation and computational efficiency. In addition, the proposed algorithms do not require any prior information and sparsity guarantee of the non-zero rows. Specifically, our algorithms still have excellent performance for the non-zero row density

channel matrix.

The remainder of this paper is organized as follows. We introduce the system model in Section II. Then, Section III presents the proposed threshold-enhanced hierarchical spatial non-stationary channel estimation. Moreover, the performance analysis of the proposed algorithms are presented in Section IV. Finally, the simulations and the conclusions are presented in Section V and Section VI, respectively.

*Notations:* $x, \mathbf{x}, \mathbf{X}$ denote scalar, vector and matrix, respectively. $\mathrm{diag}(\mathbf{x})$ is the diagonal matrix with diagonal elements $\mathbf{x}$. $\mathcal{CN}(\mu, \sigma^2)$ denotes a complex Gaussian distribution with mean $\mu$ and variance $\sigma^2$. $\mathbf{X} \otimes \mathbf{Y}$ denotes the Kronecker product of $\mathbf{X}$ and $\mathbf{Y}$. $\odot$ denotes the XNOR operation. The cardinality of $\Omega$ is defined as $|\Omega|$. $\|\cdot\|_0$, $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_F$ are $l_0$-norm, $l_1$-norm, $l_2$-norm and Frobenius-norm, respectively.

## II. System Model

Consider the spatial non-stationary channel estimation of uplink massive MIMO system, in which $K$ single-antenna users communicate with one BS equipped with $M$ antennas. The delay-domain CIRs between the $k(k \in \{1, 2, \cdots, K\})$-th user and the $m(m \in \{1, 2, \cdots, M\})$-th BS antenna can be expressed as [26] $\mathbf{h}_{k,m} = (h_{k,m1}, h_{k,m2}, \cdots, h_{k,mL})^T$, where $L$ is the equivalent channel length. Especially, only a small portion of elements in $\mathbf{h}_{k,m}$ are significant, and the remaining are near-zero or zero elements, i.e., the channel vector $\mathbf{h}_{k,m}$ can be sparsely represented in the delay-domain. Let $\bar{\mathbf{p}}_k \in \mathbb{C}^{N \times 1}$ be the transmit signal including data and pilots of user $k$ in the frequency-domain, which is further transformed into signal $\tilde{\mathbf{p}}_k \in \mathbb{C}^{N \times 1}$ in the time-domain after inverse discrete Fourier transform (IDFT) at the transmitter. Then, the time-domain received signal at the $m$-th antenna of the BS is

$$\breve{\mathbf{y}}_m = \tilde{\mathbf{p}}_k \circledast \mathbf{h}_{k,m} + \tilde{\mathbf{w}}_m = \mathcal{T} \begin{bmatrix} \mathbf{h}_{k,m} \\ \mathbf{0}_{N-L} \end{bmatrix} + \tilde{\mathbf{w}}_m, \quad (1)$$

where $\circledast$ denotes the circular convolution operation, $\tilde{\mathbf{w}}_m$ is the additive Gaussian white noise (AWGN). Especially, $\mathcal{T} \in \mathbb{C}^{N \times N}$ is the Toeplitz matrix (or circulant matrix), which is determined by the $\tilde{\mathbf{p}}_k$ and defined as

$$\mathcal{T} = \begin{bmatrix} \tilde{p}_0 & \tilde{p}_{N-1} & \cdots & \tilde{p}_1 \\ \tilde{p}_1 & \tilde{p}_0 & \cdots & \tilde{p}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_{N-1} & \tilde{p}_{N-2} & \cdots & \tilde{p}_0 \end{bmatrix}. \quad (2)$$

Because fast Fourier transform can be used to diagonalize the circulant matrix, i.e.,

$$\mathcal{T} = \mathbf{F}_{N \times N}^{-1} \mathrm{diag}(\bar{\mathbf{p}}_k) \mathbf{F}_{N \times N}, \quad (3)$$

with an $N$-point discrete Fourier transform (DFT) matrix $\mathbf{F}_{N \times N}$, the frequency-domain signal can be represented as

$$\bar{\mathbf{y}}_m = \mathbf{F}_{N \times N} \breve{\mathbf{y}}_m = \mathbf{F}_{N \times N} \mathcal{T} \begin{bmatrix} \mathbf{h}_{k,m} \\ \mathbf{0}_{N-L} \end{bmatrix} + \mathbf{F}_{N \times N} \tilde{\mathbf{w}}_m \quad (4)$$

$$= \mathrm{diag}(\bar{\mathbf{p}}_\mathbf{k}) \mathbf{F}_{N \times L} \mathbf{h}_{k,m} + \bar{\mathbf{w}}_m,$$

where $\mathbf{F}_{N \times L}$ denotes the selection of the first $L$ columns of $\mathbf{F}_{N \times N}$, $\bar{\mathbf{w}}_m = \mathbf{F}_{N \times N} \tilde{\mathbf{w}}_m$ denotes the AWGN in the

frequency-domain. Without loss of generality, we assume that pilot signals are inserted on $N_p$ subcarriers. Let $\mathbf{p}_k \in \mathbb{C}^{N_p \times 1}$ be the pilot part of transmit data $\bar{\mathbf{p}}_k$ for user $k$ in frequency-domain. Then, the received pilots in frequency-domain from user $k$ at the $m$-th BS antenna for channel estimation is expressed as

$$\mathbf{y}_m = \mathrm{diag}(\mathbf{p}_k) \mathbf{F}_{N_p \times L} \mathbf{h}_{k,m} + \mathbf{w}_m \triangleq \mathbf{A}_k \mathbf{h}_{k,m} + \mathbf{w}_m, \quad (5)$$

where $\mathbf{p}_k = (e^{j\theta_{k,1}}, e^{j\theta_{k,2}}, \cdots, e^{j\theta_{k,N_p}})^T \in \mathbb{C}^{N_p \times 1}$ with $\{\theta_{k,n}\}_{n=1}^{N_p} \in [0, 2\pi]$ is the pilot sequence of user $k$, $\mathbf{F}_{N_p \times L} \in \mathbb{C}^{N_p \times L}$ is a DFT sub-matrix, $\mathbf{A}_k = \mathrm{diag}(\mathbf{p}_k) \mathbf{F}_{N_p \times L} \in \mathbb{C}^{N_p \times L}$, and $\mathbf{w}_m \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_p}) \in \mathbb{C}^{N_p \times 1}$ is the corresponding AWGN for channel estimation.

At the $m$-th BS antenna, the received frequency-domain signals from all users are expressed as

$$\mathbf{y}_m = \sum_{k=1}^{K} \mathbf{A}_k \mathbf{h}_{k,m} + \mathbf{w}_m \triangleq \mathbf{S} \mathbf{x}_m + \mathbf{w}_m, \quad (6)$$

where $\mathbf{S} = [\mathbf{A}_1, \cdots, \mathbf{A}_K] \in \mathbb{C}^{N_p \times KL}, \mathbf{x}_m = [\mathbf{h}_{1,m}^T, \mathbf{h}_{2,m}^T, \cdots, \mathbf{h}_{K,m}^T]^T \in \mathbb{C}^{KL \times 1}$.

Further, the received signals at all BS antennas can be formulated as

$$\mathbf{Y} = \mathbf{S}\mathbf{X} + \mathbf{W}, \quad (7)$$

where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_M] \in \mathbb{C}^{N_p \times M}$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M] \in \mathbb{C}^{KL \times M}$, and $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_M] \in \mathbb{C}^{N_p \times M}$. Note that $\mathbf{X}$ is a non-uniform block sparse matrix, i.e., spatial non-stationary channel or the sparse structures of delay-domain channels for different antennas are not identical [30], in which most rows have all zero elements, and the remaining rows have some non-zero elements. Based on the sparse structure of $\mathbf{X}$, the channel estimation in (7) is transformed into solving the following $l_0$-norm minimization problem, i.e.,

$$\min_{\mathrm{vec}(\mathbf{X}) \in \mathbb{C}^{KLM \times 1}} \|\mathrm{vec}(\mathbf{X})\|_0 \quad (8a)$$

$$\text{s.t.} \quad \|\mathrm{vec}(\mathbf{Y}) - (\mathbf{I}_M \otimes \mathbf{S}) \mathrm{vec}(\mathbf{X})\|_2^2 \le \varrho, \quad (8b)$$

where $\mathrm{vec}(\mathbf{X})$ denotes matrix $\mathbf{X}$ vectorization, and $\varrho \ge 0$ is a tuning parameter introduced by measurement noise. However, (8) is an NP-hard problem. The objective function of (8) can be replaced by other norms to find the sparse solution approximatively. In particular, the $\|\mathbf{X}\|_{2,0}$ or $\|\mathbf{X}\|_{2,1}$ is further considered when $\mathbf{X}$ is a uniform block sparse matrix, i.e., the spatial stationary channel. The goal of this paper is to design an efficient algorithm for solving the problem (8) with non-uniform block sparse matrix $\mathbf{X}$.

## III. Threshold-Enhanced Hierarchical Spatial Non-Stationary Channel Estimation

In this section, we propose a threshold-enhanced hierarchical algorithm to estimate the non-uniform block sparse channel matrix, in which the non-zero rows are estimated in the first layer and the elements of the estimated non-zero rows are further estimated in the second layer. In particular, a row-wise SAMP algorithm is proposed for estimating the non-uniform

block sparse channel matrix. To further estimate the positions of zero elements in non-zero rows, a threshold is designed in the second layer.

### A. Threshold-Enhanced Hierarchical Algorithm

*1) Row-Wise Estimation Layer:* The estimated non-uniform block sparse matrix $\hat{\mathbf{X}}^{(t)}$ of problem (8) in the $t$-th iteration can be expressed as

$$\hat{\mathbf{X}}^{(t)} = [(\hat{\mathbf{x}}_1^{(t)})^T, (\hat{\mathbf{x}}_2^{(t)})^T, \cdots, (\hat{\mathbf{x}}_{KL}^{(t)})^T]^T, \tag{9}$$

where $\hat{\mathbf{x}}_r^{(t)}, r = 1, 2, \cdots, KL$ is the $r$-th row of matrix $\hat{\mathbf{X}}^{(t)}$. Defining $\Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})$ as the row support set of $\hat{\mathbf{X}}^{(t)}$, it is clear that $\Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)}) = \{r | \|\hat{\mathbf{x}}_r^{(t)}\|_2^2 \neq 0, r = 1, 2 \cdots, KL\}$. Note that $\hat{\mathbf{X}}^{(t)}$ has many zero rows, which can be estimated by using the SAMP algorithm.

Different from the element-wise SAMP estimation, we propose a row-wise SAMP algorithm for the first layer estimation, in which the in-row correlation of the non-uniform sparse channel matrix is considered to improve the estimation accuracy. The purpose of this layer is to find the position of the non-zero rows in $\hat{\mathbf{X}}^{(t)}$ quickly and estimate the corresponding elements. To this end, a candidate row support set, defined as $\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})$, is first introduced. The main processing includes the following two steps:

**Step 1**: Update the candidate row support set $\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})$. In order to the row support set $\Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})$, the sensing matrix can be expressed as $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_{KL}]$, where $\mathbf{s}_r$ is the $r$-th column of $\mathbf{S}$. The correlation coefficient between each column of $\mathbf{S}$ and the residual matrix $\mathbf{R}^{(t-1)}$ of the $(t-1)$-th iteration is defined as

$$c_r^{(t)} = \|\mathbf{s}_r^H \mathbf{R}^{(t-1)}\|_1, r = 1, 2, \cdots, KL, \tag{10}$$

where $\mathbf{R}^{(t-1)} = \mathbf{Y} - \mathbf{S}_{\Omega_{\text{row}}^{(t-1)}(\hat{\mathbf{X}}^{(t-1)})} \hat{\mathbf{X}}_{\Omega_{\text{row}}^{(t-1)}}^{(t-1)} \in \mathbb{C}^{N_p \times M}$ is the residual matrix with $\mathbf{R}^{(0)} = \mathbf{Y}$, and $\mathbf{S}_{\Omega_{\text{row}}^{(t-1)}(\hat{\mathbf{X}}^{(t-1)})} \in \mathbb{C}^{N_p \times |\Omega_{\text{row}}^{(t-1)}(\hat{\mathbf{X}}^{(t-1)})|}$ is the sub-matrix containing the columns selected from $\mathbf{S}$ according to the row support set $\Omega_{\text{row}}^{(t-1)}(\hat{\mathbf{X}}^{(t-1)})$. The larger value of $c_r^{(t)}$ means that $\mathbf{s}_r$ is more correlated with $\mathbf{R}^{(t-1)}$, and the $r$-th row of $\mathbf{X}$ is more likely to be a non-zero row. Defining the index set as $\Pi^{(t)} = \{r | c_r^{(t)}, r = 1, 2, \cdots, KL\}$, then it can be split into two subsets, i.e.,

$$\Pi^{(t)} = \Lambda_I^{(t)} \cup \Lambda_{KL-I}^{(t)}, \tag{11}$$

where $\Lambda_I^{(t)} \cap \Lambda_{KL-I}^{(t)} = \varnothing, |\Lambda_I^{(t)}| = I, |\Lambda_{KL-I}^{(t)}| = KL - I$, and $I$ is the size of preliminary candidate row support set. In particular, the defined subsets $\Lambda_I^{(t)}$ and $\Lambda_{KL-I}^{(t)}$ satisfy the following condition:

$$\forall j, k, j \in \Lambda_I^{(t)}, k \in \Lambda_{KL-I}^{(t)}, c_j^{(t)} \geq c_k^{(t)}, \tag{12}$$

i.e., $\Lambda_I^{(t)}$ contains the indexes of the $I$ maximum values of $\{c_r^{(t)}\}_{r=1}^{KL}$.

The update of candidate row support set:

$$\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)}) = \Omega_{\text{row}}^{(t-1)}(\hat{\mathbf{X}}^{(t-1)}) \cup \Lambda_I^{(t)}, \tag{13}$$

where $\Omega_{\text{row}}^{(0)}(\hat{\mathbf{X}}^{(0)}) = \varnothing$.

**Step 2**: Update the row support set $\Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})$. With the candidate set $\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})$ updated in step 1, the estimated channel matrix $\hat{\mathbf{X}}^{(t)}$ is

$$\hat{\mathbf{X}}^{(t)} = \mathbf{S}_{\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})}^{\dagger} \mathbf{Y}, \tag{14}$$

where $\mathbf{S}_{\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})}^{\dagger} \in \mathbb{C}^{|\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})| \times N_p}$ is the pseudo-inverse of $\mathbf{S}_{\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})}$. To quickly find non-zero rows, the $\ell_2$-norm of each row of $\hat{\mathbf{X}}^{(t)}$ is calculated as

$$f_r^{(t)} = \|\hat{\mathbf{x}}_r^{(t)}\|_2^2, r = 1, 2, \cdots, KL. \tag{15}$$

Note that if $r \notin \Psi_{row}^{(t)}(\hat{\mathbf{X}}^{(t)})$, $\hat{\mathbf{x}}_r^{(t)} = \mathbf{0}$. Similar to (13), the $\Xi^{(t)} = \{r | f_r^{(t)}, r = 1, 2, \cdots, KL\}$ can be expressed as

$$\Xi^{(t)} = \Delta_I^{(t)} \cup \Delta_{KL-I}^{(t)}, \tag{16}$$

where $\Delta_I^{(t)} \cap \Delta_{KL-I}^{(t)} = \varnothing, |\Delta_I^{(t)}| = I, |\Delta_{KL-I}^{(t)}| = KL - I$, and the defined subsets $\Delta_I^{(t)}$ and $\Delta_{KL-I}^{(t)}$ satisfy

$$\forall j, k, j \in \Delta_I^{(t)}, k \in \Delta_{KL-I}^{(t)}, f_j^{(t)} \geq f_k^{(t)}. \tag{17}$$

Then the row support set is estimated as $\Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)}) = \Delta_I^{(t)}$.

Update the candidate set and the row support set in the first layer until the residual matrix satisfies the halting condition.

Finally, the non-zero rows can be found, and the channel matrix is estimated by

$$\hat{\mathbf{X}}_{\text{1st}} = \mathbf{S}_{\Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})}^{\dagger} \mathbf{Y} \in \mathbb{C}^{|\Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})| \times M}. \tag{18}$$

**Remark 1.** *The $\ell_2$-norm of each row of $\hat{\mathbf{X}}^{(t)}$ is calculated in (15), and the $I$ rows with larger $l_2$-norm are found according to (16) and (17). We are actually calculating the $l_{2,0}$-norm $\|\mathbf{X}\|_{2,0}^2$, which can improve the efficiency of finding the row support set of the channel matrix.*

**Remark 2.** *The zero elements of $\mathbf{X}$ are estimated according to the complement of row support set $\Omega_{\text{row}}(\hat{\mathbf{X}}_{\text{1st}})$, i.e., all non-zero rows are found; while the zero elements in non-zero rows can not be estimated.*

It is important to point out that the non-zero rows are always found perfectly due to the traversal of all possibilities, and the values of non-zero rows and zero rows are separated in $\Xi^{(t)}$. However, the estimation accuracy of the non-uniform sparse channel matrix in (18) is affected by the sparsity of the non-zero row. Specifically, it has high estimation accuracy if the non-zero row of the channel matrix has large sparsity, i.e., a large number of non-zero elements, especially for the case that the non-zero row is density, such as the stationary channel. To estimate the elements of non-zero row without sparsity prior information, an element-wise estimation for the estimated non-zero rows is proposed in the second layer.

*2) Element-Wise Estimation Layer:* Let $\mathbb{E} \triangleq \Omega_{\text{row}}(\hat{\mathbf{X}}_{\text{1st}})$, $\mathbf{S}_{\mathbb{E}}$ denotes the matrix composed of the corresponding columns of $\mathbb{E}$ in matrix $\mathbf{S}$, and $\mathbf{X}^{(\mathbb{E})}$ denotes the matrix composed of the corresponding rows of $\mathbb{E}$ in the matrix $\hat{\mathbf{X}}_{\text{1st}}$, then the received signal in (7) is represented as:

$$\mathbf{Y} = \mathbf{S}_{\mathbb{E}} \mathbf{X}^{(\mathbb{E})} + \mathbf{W}. \tag{19}$$

Note that the column vectors of $\mathbf{S}_{\mathbb{E}}$ estimated by row-wise SAMP in the first layer are linearly independent, i.e., $\mathbf{S}_{\mathbb{E}}$ is a column nonsingular matrix. Further, the received signal can be re-written as

$$\tilde{\mathbf{Y}} = \mathbf{X}^{(\mathbb{E})} + \tilde{\mathbf{W}}, \tag{20}$$

where $\bar{\mathbf{S}}_{\mathbb{E}} = (\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})^{-1}\mathbf{S}_{\mathbb{E}}^H$, $\tilde{\mathbf{Y}} \triangleq \bar{\mathbf{S}}\mathbf{Y}, \tilde{\mathbf{W}} \triangleq \bar{\mathbf{S}}\mathbf{W}$. Therefore, the regularization problem for estimating $\mathbf{X}^{(\mathbb{E})}$ in the second layer of the proposed algorithm is formulated as

$$\hat{\mathbf{X}}^{(\mathbb{E})} = \arg \min_{\mathbf{X}^{(\mathbb{E})}\in\mathbb{C}^{|\mathbb{E}|\times M}} \frac{1}{2}\|\tilde{\mathbf{Y}} - \mathbf{X}^{(\mathbb{E})}\|_F^2 + \zeta\|\mathbf{X}^{(\mathbb{E})}\|_{0,1}, \tag{21}$$

where $\zeta > 0$ is a regular parameter, and $\|\mathbf{X}^{(\mathbb{E})}\|_{0,1} \triangleq \sum_{c=1}^{M}\|\mathbf{x}_{\mathbb{E},c}\|_0$ is $\ell_{0,1}$-norm encouraging the columns of matrix $\mathbf{X}^{(\mathbb{E})}$ to become sparse, where $\mathbf{x}_{\mathbb{E},c}$ is the $c$-th column of $\mathbf{X}^{(\mathbb{E})}$. Further, (21) is expressed as

$$\{\hat{\mathbf{x}}_{\mathbb{E},c}\}_{c=1}^{M} = \arg \min_{\{\mathbf{x}_{\mathbb{E},c}\}_{c=1}^{M}} \sum_{c=1}^{M}\left(\frac{1}{2}\|\tilde{\mathbf{y}}_c - \mathbf{x}_{\mathbb{E},c}\|_2^2 + \zeta\|\mathbf{x}_{\mathbb{E},c}\|_0\right), \tag{22}$$

where $\tilde{\mathbf{y}}_c$ is the $c$-th column of $\tilde{\mathbf{Y}}$. The closed-form of the solution in problem (22) can be obtained by the following Theorem.

**Theorem 1.** *The element of non-zero rows in the channel matrix is given by*

$$\hat{x}_{\mathbb{E},cn} = \begin{cases} 0, & if \quad \zeta_0 \geq |\tilde{y}_{cn}|^2, \\ \tilde{y}_{cn}, & if \quad 0 < \zeta_0 < |\tilde{y}_{cn}|^2, \end{cases} \tag{23}$$

*and the threshold $\zeta_0$ is expressed as*

$$\zeta_0 = \|(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})^{-1}\mathbf{S}_{\mathbb{E}}^H\|_2^2 \frac{N_p \sigma^2}{|\mathbb{E}|}\kappa, \tag{24}$$

*where $\hat{x}_{\mathbb{E},cn}$ and $\tilde{y}_{cn}$ represent the $n$-th element of $\hat{\mathbf{x}}_{\mathbb{E},c}$ and $\tilde{\mathbf{y}}_c$, respectively. In addition, $\kappa \in [0,1]$ is a reconstructed noise level parameter.*

*Proof:* Please refer to Appendix A. ∎

Hence, the proposed threshold-enhanced hierarchical algorithm can be described in the Algorithm 1. It is important to note that the step size of preliminary candidate row support set is 1, and one time least squares estimation is performed in each iteration. The computational complexity varies with the scale of the communication system. To overcome this challenge, we next propose an adaptive threshold-enhanced-hierarchical Algorithm.

### B. Adaptive Threshold-Enhanced Hierarchical Algorithm

Let $\Upsilon = [1, 2, ..., KL]$ be the index set of matrix $\mathbf{S}$, the support set $\hat{\mathbb{E}}$ which has been selected and the complement $\hat{\mathbb{E}}^C$ which has not been selected. Suppose the correct support set is $\mathbb{E}$. By increasing the step size of preliminary candidate row support set and the candidate set size in Algorithm 1, we can find an approximate support set as quickly as possible. Specifically, in Step 1, we can add more than one row at each time to update the candidate row support set $\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})$, i.e.,

---

**Algorithm 1** Proposed TEHE Algorithm

**Input:** Matrices $\mathbf{Y}$ and $\mathbf{S}$.
**Output:** Estimated channel matrix $\hat{\mathbf{X}}$.

1: **Row-Wise Estimation Layer:**
2: **Initialization:** $I = 1, t = 1, \Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(0)}) = \emptyset, \mathbf{R}^{(0)} = \mathbf{Y}$.
3: **Repeat**
4: $\quad \Lambda_I^{(t)} = \max\{[c_r]_{r=1}^{KL}, I\}$; (Preliminary Set)
5: $\quad \Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)}) = \Omega_{\text{row}}^{(t-1)}(\hat{\mathbf{X}}^{(t-1)}) \cup \Lambda_I^{(t)}$; (Candidate Set)
6: $\quad \Delta_I^{(t)} = \max\{[\|\mathbf{S}_{\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})_r}^\dagger \mathbf{Y}\|_2^2]_{r=1}^{|\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})|}, I\}$; (Final Test)
7: $\quad \hat{\mathbf{X}}_{1st}^{(t)} = \mathbf{S}_{(\Delta_I^{(t)})}^\dagger \mathbf{Y}$; (Least Squares Estimation)
8: $\quad \mathbf{R}_{(\Delta_I^{(t)})} = \mathbf{Y} - \mathbf{S}_{\Delta_I^{(t)}}\hat{\mathbf{X}}_{1st}^{(t)}$; (Residual Calculation)
9: $\quad$ **if** the halting condition is met **then**
10: $\qquad$ **break;**
11: $\quad$ **else**
12: $\qquad$ **if** $\|\mathbf{R}_{(\Delta_I^{(t)})}\|_2 \geq \|\mathbf{R}^{(t-1)}\|_2$ **then**
13: $\qquad\quad I = I + 1$; (Set Size Increase)
14: $\qquad$ **else**
15: $\qquad\quad \Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)}) = \Delta_I^{(t)}; \mathbf{R}^{(t)} = \mathbf{R}_{(\Delta_I^{(t)})}; t = t + 1$; (Update)
16: $\qquad$ **end if**
17: $\quad$ **end if**
18: **Element-Wise Estimation Layer:**
19: Refine element according to (40) in Theorem 1
20: **return** result

---

let the step size $s > 1$, and $s$ can be selected as large as possible.

However, if the step size is increased there will be extra non-zero rows in the selected support set when the first layer stop, which satisfies $\hat{\mathbb{E}} \supseteq \mathbb{E}$. Thus, it is necessary to remove these extra rows. It can be known that the received signal of the $(t)$-th iteration can be expressed as

$$\begin{aligned} &(\mathbf{S}_{\hat{\mathbb{E}}}^H \mathbf{S}_{\hat{\mathbb{E}}})^{-1}\mathbf{S}_{\hat{\mathbb{E}}}^H \mathbf{Y} \\ &= (\mathbf{S}_{\hat{\mathbb{E}}}^H \mathbf{S}_{\hat{\mathbb{E}}})^{-1}\mathbf{S}_{\hat{\mathbb{E}}}^H \mathbf{S}_{\hat{\mathbb{E}}}\mathbf{X}^{(\hat{\mathbb{E}})} + (\mathbf{S}_{\hat{\mathbb{E}}}^H \mathbf{S}_{\hat{\mathbb{E}}})^{-1}\mathbf{S}_{\hat{\mathbb{E}}}^H \mathbf{S}_{\hat{\mathbb{E}}^C}\mathbf{X}^{(\hat{\mathbb{E}}^C)} \\ &\quad + (\mathbf{S}_{\hat{\mathbb{E}}}^H \mathbf{S}_{\hat{\mathbb{E}}})^{-1}\mathbf{S}_{\hat{\mathbb{E}}}^H \mathbf{W}, \end{aligned} \tag{25}$$

where $\mathbf{X}^{(\hat{\mathbb{E}})}$ is a submatrix composed of the rows of $\mathbf{X}$ indexed by $\hat{\mathbb{E}}$. Note that $\hat{\mathbb{E}} \supseteq \mathbb{E}$ if the iteration stops, then $\hat{\mathbb{E}}^C$ does not belong to the real support set. As a result, $\mathbf{X}^{(\hat{\mathbb{E}}^C)} = 0$, and the received signal can be re-written as

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}} + \tilde{\mathbf{W}}, \tag{26}$$

where $\tilde{\mathbf{Y}} \triangleq \mathbf{S}_{\hat{\mathbb{E}}}^\dagger \mathbf{Y}, \tilde{\mathbf{X}} \triangleq \mathbf{X}^{(\hat{\mathbb{E}})}, \tilde{\mathbf{W}} \triangleq \mathbf{S}_{\hat{\mathbb{E}}}^\dagger \mathbf{W}$, and $\mathbf{S}_{\hat{\mathbb{E}}}^\dagger = (\mathbf{S}_{\hat{\mathbb{E}}}^H \mathbf{S}_{\hat{\mathbb{E}}})^{-1}\mathbf{S}_{\hat{\mathbb{E}}}$. Then, we should find out the correct support set $\mathbb{E}$ from the estimated support set $\hat{\mathbb{E}}$. In particular, the regularization problem for estimating $\mathbf{X}$ in the row-wise estimation layer which is to find the correct support set $\mathbb{E}$ can be formulated as

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}\in\mathbb{C}^{|\hat{\mathbb{E}}|\times M}} \frac{1}{2}\|\tilde{\mathbf{Y}} - \mathbf{X}\|_F^2 + \eta\|\mathbf{X}^T\|_{2,0}, \tag{27}$$

where $\eta > 0$ is a regular parameter, and $\|\mathbf{X}\|_{2,0} = \sum_{r=1}^{|\hat{\mathbb{E}}|}\mathbb{I}_{\|\mathbf{x}_r\|_2}$ is $\ell_{2,0}$-norm encouraging the rows of matrix $\mathbf{X}$

to be zeros, where $\mathbb{I}_{\|\mathbf{x}_r\|_2} = 1$ for $\|\mathbf{x}_r\|_2 \neq 0$, otherwise, $\mathbb{I}_{\|\mathbf{x}_r\|_2} = 0$. Further, (27) is expressed as

$$\{\hat{\mathbf{x}}_r\}_{r=1}^{|\hat{\mathbb{E}}|} = \arg\min_{\{\mathbf{x}_r\}_{r=1}^{|\hat{\mathbb{E}}|}} \sum_{r=1}^{|\hat{\mathbb{E}}|} \left( \frac{1}{2} \|\tilde{\mathbf{y}}_r - \mathbf{x}_r\|_2^2 + \eta \mathbb{I}_{\|\mathbf{x}_r\|_2 \neq 0} \right). \quad (28)$$

The closed-form solution to the problem (28) can be obtained by the following corollary which is similar to the result of Theorem 1.

**Corollary 1.** *The element in the matrix $\tilde{\mathbf{X}}$ is given by*

$$\hat{x}_{r,n} = \begin{cases} 0, & if \quad \eta_0 \geq |\tilde{y}_{r,n}|^2, \\ \tilde{y}_{r,n}, & if \quad 0 < \eta_0 < |\tilde{y}_{r,n}|^2, \end{cases} \quad (29)$$

*where the threshold $\eta_0$ is expressed as*

$$\eta_0 = \|(\mathbf{S}_\mathbb{E}^H \mathbf{S}_\mathbb{E})^{-1} \mathbf{S}_\mathbb{E}^H\|_2^2 \frac{N_p \sigma^2}{|\mathbb{E}|} \kappa, \quad (30)$$

*where $\hat{x}_{r,n}$ and $\tilde{y}_{r,n}$ represent the $n$-th element of $\hat{\mathbf{x}}_r$ and $\tilde{\mathbf{y}}_r$, respectively. In addition, $\kappa \in [0,1]$ is a reconstructed noise level parameter.*

*Proof:* Please refer to Appendix B. ∎

**Remark 3.** *Although the forms of the solution of Theorem 1 and Corollary 1 are the same, they have different roles for channel estimations. More precisely, the result of Theorem 1 is used to refine the elements in the non-zero rows, while the result of Corollary 1 is employed to find both zero rows and elements of non-zero rows in the matrix $\tilde{\mathbf{X}}$.*

The specific process is shown in Algorithm 2. With the support of Corollary 1, the step size in Algorithm 1 can be directly increased and speed up the algorithm without any other operations.

## IV. PERFORMANCE ANALYSIS

In this section, we provide theoretical guarantees that the proposed algorithms can accurately estimate the rows of channel matrix in the first layer. In addition, the halting conditions of the first layer of the proposed algorithms are analyzed and discussed. Finally, the computation complexity is analyze to evaluate the computational efficiency.

### A. Accurate Recovery

Note that (7) can be reformulated as

$$\mathbf{Y} = \mathbf{S}\mathbf{X}_{(k)} + \mathbf{W}, \quad (31)$$

for the first layer estimation, where $\mathbf{X}_{(k)}$ means that $\mathbf{X}$ has $k$ non-zero rows. To further prove sufficient conditions for precise sparse recovery in (31), we introduce the following definitions.

**Definition 1.** *Restricted Isometry Property (RIP) [21], [22]. For the $k$-sparse signal vector $\tilde{\mathbf{x}}$, the isometry constant $\delta_k$ of a matrix $\mathbf{S}$ as the smallest number such that*

$$(1 - \delta_k)\|\tilde{\mathbf{x}}\|_2^2 \leq \|\mathbf{S}\tilde{\mathbf{x}}\|_2^2 \leq (1 + \delta_k)\|\tilde{\mathbf{x}}\|_2^2 \quad (32)$$

---

**Algorithm 2** Proposed A-TEHE Algorithm

**Input:** $\mathbf{Y}$, $\mathbf{S}$, and $s$.
**Output:** Estimated channel matrix $\hat{\mathbf{X}}$.
1: **Row-Wise Estimation Layer:**
2: **Initialization:** $I = s$, $t = 1$, $\Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(0)}) = \emptyset$, $\mathbf{R}^{(0)} = \mathbf{Y}$.
3: **Repeat**
4: $\quad \Lambda_I^{(t)} = \max\{[c_r]_{r=1}^{KL}, I\}$; (Preliminary Set)
5: $\quad \Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)}) = \Omega_{\text{row}}^{(t-1)}(\hat{\mathbf{X}}^{(t-1)}) \cup \Lambda_I^{(t)}$; (Candidate Set)
6: $\quad \Delta_I^{(t)} = \max\{[\|\mathbf{S}_{\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})_r}^\dagger \mathbf{Y}\|_2^2]_{r=1}^{|\Psi_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)})|}, I\}$; (Final Test)
7: $\quad \hat{\mathbf{X}}_{1st}^{(t)} = \mathbf{S}_{\Delta_I^{(t)}}^\dagger \mathbf{Y}$; (Least Squares Estimation)
8: $\quad \mathbf{R}_{(\Delta_I^{(t)})} = \mathbf{Y} - \mathbf{S}_{\Delta_I^{(t)}} \hat{\mathbf{X}}_{1st}$; (Residual Calculation)
9: **if** halting condition is met **then**
10: $\quad$ **break;**
11: **else**
12: $\quad$ **if** $\|\mathbf{R}_{(\Delta_I^{(t)})}\|_2 \geq \|\mathbf{R}^{(t-1)}\|_2$ **then**
13: $\quad\quad I = I + s$; (Set Size Increase)
14: $\quad$ **else**
15: $\quad\quad \Omega_{\text{row}}^{(t)}(\hat{\mathbf{X}}^{(t)}) = \Delta_I^{(t)}$; $\mathbf{R}^{(t)} = \mathbf{R}_{(\Delta_I^{(t)})}$; $t = t + 1$; (Update)
16: $\quad$ **end if**
17: **end if**
18: **Element-Wise Estimation Layer:**
19: Refine element according to according to Corollary 1
20: **return** result

---

holds.

**Definition 2.** *Exact Recovery Condition (ERC) [23]. Let $\mathbf{S}_\mathbb{E}$ be the set of columns of $\mathbf{S}$ corresponding to the support set $\mathbb{E}$, define*

$$\text{ERC}(\mathbb{E}) = 1 - \max_{\mathbf{s} \in \mathbf{S}_{\mathbb{E}C}} \{\|(\mathbf{S}_\mathbb{E}^H \mathbf{S}_\mathbb{E})^{-1} \mathbf{S}_\mathbb{E}^H \mathbf{s}\|_1\}. \quad (33)$$

*Then, condition $\text{ERC}(\mathbb{E}) > 0$ is called ERC.*

**Definition 3.** *Mutual Incoherence Property (MIP) [23]. The mutual incoherence is defined by*

$$\mu = \max_{i \neq j} | < \mathbf{s}_i, \mathbf{s}_j > |, \quad (34)$$

*where $< \mathbf{s}_i, \mathbf{s}_j >$ represents the inner product of column vectors $\mathbf{s}_i$ and $\mathbf{s}_j$ in matrix $\mathbf{S}$.*

To prove that the proposed algorithm can find the non-zero row support set exactly, we need to use the following lemmas:

**Lemma 1.** *[21] if the MIP $\mu < \frac{1}{2k-1}$, then the $\text{ERC}(\mathbb{E}) \geq \frac{1-(2k-1)\mu}{1-(k-1)\mu} > 0$. where $\mu$ is the mutual incoherence of matrix $\mathbf{S}$.*

**Remark 4.** *Lemma 1 shown that if $\mu < \frac{1}{2k-1}$ in matrix $\mathbf{S}$, then $\mathbf{S}$ must satisfy ERC. In fact, the MIP condition is stricter than RIP and ERC, i.e., MIP holds implies that RIP and ERC hold but the converse is not true.*

**Lemma 2.** *Let $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ represent the minimum and maximum singular values of a matrix, respectively, we*

*have the following results [23]:*

$$\lambda_{\min}(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}}) \leq \lambda_{\min}(\mathbf{P}) \leq \lambda_{\max}(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}}) \leq \lambda_{\max}(\mathbf{P}), \quad (35)$$

$$1 - (k-1)\mu \leq \lambda_{\min}(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}}) \leq \lambda_{\max}(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})$$
$$\leq 1 + (k-1)\mu, \text{if } \mu < \frac{1}{k-1}, \quad (36)$$

*where* $\mathbf{P} = \mathbf{S}_{\mathbb{E}'_{(t)}}^H (\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t)}} \mathbf{S}_{\mathbb{E}_{(t)}}^\dagger) \mathbf{S}_{\mathbb{E}'_{(t)}}$, *and* $\mathbb{E}'_{(t)} = \mathbb{E} - \mathbb{E}_{(t)}$, *i.e.,* $\mathbb{E}'_{(t)} \bigcup \mathbb{E}_{(t)} = \mathbb{E}$ *and* $\mathbb{E}'_{(t)} \bigcap \mathbb{E}_{(t)} = \varnothing$.

From these we can derive a sufficient condition for exact recovery, which is as shown in the following theorem,

**Theorem 2.** *Suppose* $\mu < \frac{1}{2k-1}$, *and all* $l_2$-*norm of nonzero rows* $\|\hat{\bar{\mathbf{x}}}^{(i)}\|_2$ *satisfy* $\|\hat{\bar{\mathbf{x}}}^{(i)}\|_2 \geq \frac{2\sigma\sqrt{N_p M}}{1-(2k-1)\mu}$, *the proposed Algorithm 1 can recover exactly the true support set* $\mathbb{E}$ *and the vector* $\hat{\bar{\mathbf{x}}}$ *with a suitable halting condition.*

*Proof:* Please refer to Appendix C. ∎

Theorem 2 states that the Row-Wise Estimation Layer in Algorithm 1 can accurately estimate all correct support sets as long as two conditions are satisfied. The sufficient condition $\mu < \frac{1}{2k-1}$ is a constraint on pilot matrix $\mathbf{S}$. It requires $\mathbf{S}$ to satisfy the general compressed sensing accurate recovery condition, which can be guaranteed by the design of the pilot. For another sufficient condition $\|\hat{\bar{\mathbf{x}}}^{(i)}\|_2 \geq \frac{2\sigma\sqrt{N_p M}}{1-(2k-1)\mu}$, it shows that due to the influence of noise, the power any of the remaining rows in $\mathbf{X}$ need to be large enough, the Row-Wise Estimation Layer can select correct support set at this step.

### B. Halting Condition

It can be seen from the above analysis that Algorithm 1 can gradually find all correct non-zero rows. In order for the algorithm to stop in time when all non-zero rows are found instead of continuing, we set a suitable stopping condition by following theorem:

**Theorem 3.** *The halting condition for the proposed TEHE and A-TEHE is that the power of residual* $\|\mathbf{R}\|_F^2 \leq N_p M \sigma^2$.

*Proof: Please refer to Appendix D.* ∎

### C. Complexity Analysis

In the following, we analyze the complexity of the proposed algorithm by flops (i.e., the total numbers of addition, subtraction, multiplication, and division). Compared to calculating the 2-norm and sorting, the main computational effort of the algorithm is spent on matrix pseudo-inverse and matrix multiplication. Consequently, only the complexity of matrix operations is considered below.

We analyze the complexity of each main operation first. The operation which calculates the correlation coefficient between each column of $\mathbf{S}$ and the residual matrix $\mathbf{R}^{(t-1)}$ costs $\mathcal{O}((KL)N_p M)$ flops at the $(t-1)$-th update. Then validation of candidate support sets and least squares operations cost

$\mathcal{O}(|\mathbb{E}|^2 N_p + |\mathbb{E}| N_p M)$ flops. For the Proposed-TEHE, it needs to go through $|\mathbb{E}|$ cycles, the complexity of the Proposed-TEHE is $\mathcal{O}(|\mathbb{E}|(KL)N_p M) + \mathcal{O}(|\mathbb{E}|^3 N_p) + \mathcal{O}(|\mathbb{E}|^2 N_p M)$. Generally, $KL > N_p > |\mathbb{E}|$, the complexity of the Proposed-TEHE is $\mathcal{O}(|\mathbb{E}|(KL)N_p M)$. For the Proposed-ATEHE, it goes through fewer cycles than the Proposed-TEHE if the step size is set as $s > 1$, and its complexity is $\mathcal{O}(\frac{|\mathbb{E}|}{s}(KL)N_p M)$. If the conventional SAMP i.e., the Parallel-SAMP [31] algorithm is used to estimate the channel, where each antenna uses the SAMP to process the received signal independently. The complexity of one cycle is $\mathcal{O}((KL)N_p + |\mathbb{E}|^2 N_p + |\mathbb{E}| N_p M)$. Therefore, the total complexity of this algorithm is $\mathcal{O}(\mathbb{E}(KL)N_p M + |\mathbb{E}|^3 N_p M + |\mathbb{E}|^2 N_p M^2)$. In addition, we compared two other algorithms, the two-stage alternating direction method of multipliers (Two-Stage-ADMM) [32] and multiple measurement vectors-approximate message passing (MMV-AMP) [33], [34]. The complexity of these algorithms is shown as follows. The complexity of the Two-Stage-ADMM is $\mathcal{O}(T_1((KL)^2(N_p + M) + M(KL)N_p) + T_2(|\mathbb{E}|^2(N_p + M) + |\mathbb{E}| M N_p))$, while the complexity of MMV-AMP is $\mathcal{O}(T N_p(KL)M)$, where $T_1, T_2, T$ are the number of iterations. The complexity of each algorithm is detailed in Table I. It can be observed that the complexity of our proposed algorithm is lower than other algorithms.

## V. SIMULATION RESULTS

In this section, simulations are carried out to evaluate the performance of the proposed algorithms for uplink spatial non-stationary channel estimation of massive MIMO systems. In order to provide a baseline for comparison, we consider the Oracle-LS by assuming the true support set of the channel matrix known at the BS. Moreover, the performance of Parallel-SAMP [31], Tow-Stage ADMM [32], and MMV-AMP [33], [34] for channel estimation is also compared. The common parameters are set as: the number of the user $K = 8$, the equivalent channel length $L = 64$, the length of pilot $N_p = 256$, and the number of the receiving antenna $M = 128$.

### A. Performance Comparison of Different Algorithms

Firstly, we evaluate the NMSE of the channel estimation, which is calculated as

$$\mathbf{NMSE} = \mathbb{E}\left[\frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_F^2}{\|\mathbf{X}\|_F^2}\right], \quad (37)$$

where $\hat{\mathbf{X}}$ is the estimated channel.

In Fig. 2, the NMSE performance of the Parallel-SAMP, the proposed TEHE, the proposed A-TEHE, and the Oracle-LS for the uplink spatial non-stationary channel are compared. It can be observed that the NMSE performance of the proposed TEHE is slightly lower than that of Parallel-SAMP for SNR≤15dB. This can be explained as the noise level is high and the channel information does not satisfy the sufficient condition in Theorem 2. As a result, the proposed TEHE fails to find all correct non-zero rows; while the A-TEHE outperforms the Parallel-SAMP for the reason that increasing an appropriate step size will find a redundant support set

TABLE I
THE COMPLEXITY OF ALGORITHMS

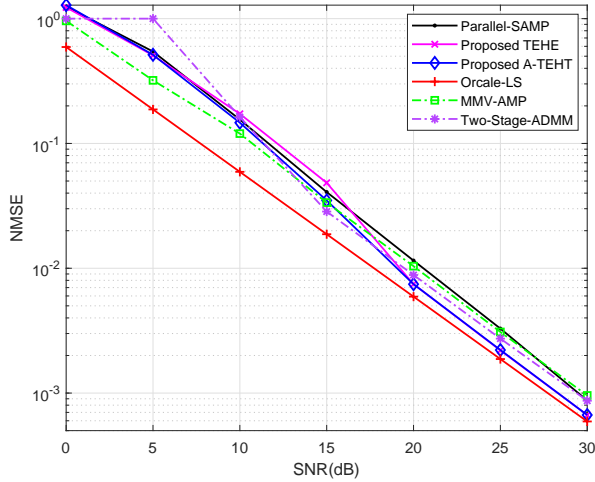| Algorithm | Complexity |
|---|---|
| Proposed-TEHE | $\mathcal{O}(|\mathbb{E}|(KL)N_pM)$ |
| Proposed-ATEHE | $\mathcal{O}(\frac{|\mathbb{E}|}{s}(KL)N_pM)$ |
| Parallel-SAMP | $\mathcal{O}(|\mathbb{E}|(KL)N_pM + |\mathbb{E}|^3N_pM + |\mathbb{E}|^2N_pM^2)$ |
| Two-Stage-ADMM | $\mathcal{O}(T_1((KL)^2(N_p + M) + M(KL)N_p) + T_2(|\mathbb{E}|^2(N_p + M) + |\mathbb{E}|MN_p))$ |
| MMV-AMP | $\mathcal{O}(TN_p(KL)M)$ |



Fig. 2. NMSE performance comparison of different non-stationary channel estimation algorithms.
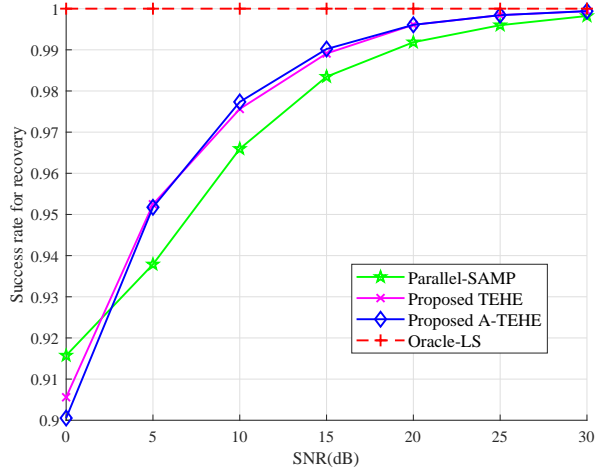


Fig. 3. Success rate for recovery with different channel estimation algorithms.

thus compensating for the disadvantages of TEHE. For SNR $\geq$ 20dB, the performance of TEHE and A-TEHE is better than Parallel-SAMP and very close to that of Oracle-LS. Our proposed algorithm underperforms MMV-AMP at low SNRs (SNR $<$ 15dB). However, to guarantee performance, the MMV-AMP need to utilize the prior information of the channel. As the SNR increases, MMV-AMP begins to underperform the other algorithms and the NMSE performance of our proposed algorithm begins to emerge. When SNR $>$ 15, our algorithm consistently outperforms MMV-AMP and Two-Stage ADMM.

In addition, the success rate for recovery is regarded as another performance evaluation metric of channel estimation, which is calculated as $\frac{\sum_{i=1}^{KL}\sum_{j=1}^{M}(\hat{x}_{i,j}\odot x_{i,j})}{KL*M}$. It indicates the proportion of zero and non-zero locations correctly estimated by the recovery algorithm. Fig. 3 demonstrates the success rates for recovery with different algorithms. It is obvious that the success rate for recovery increases as the SNR is larger. Especially, the success rate for recovery is close to $100\%$ after SNR $=$ 30dB. Besides, the successful recovery rates of the two proposed algorithms are consistently higher than that of Parallel-SAMP. This shows the advantages of the proposed algorithm in finding the correct support set. Therefore, using the non-uniform sparse feature of spatial non-stationary channel matrix, not only the complexity of the channel estimation algorithm can be reduced but also the accuracy can be improved.
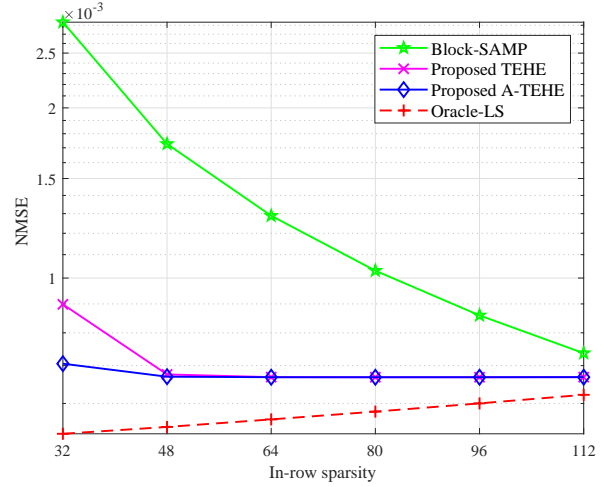
### B. Performance Comparison of Different Methods in The Second Layer.

Fig. 4 (a) compares the NMSE performance for different algorithms with different processing methods in the second layer. Especially, the block-SAMP here only has one layer for channel estimation. It can be clearly seen that the NMSE performance is improved by using threshold processing in the second layer. In fact, the block-SAMP is suitable for spatial stationary channels for the reason that it can estimate the non-zero rows correctly and all the elements in non-zero rows are non-zero. However, in the spatial non-stationary channel, there are zero entities in the non-zero rows. Therefore, the performance of spatial stationary channel estimation algorithm decreases for the case that the non-zero rows are sparse. If the second layer is added using a threshold approach, the noise in the first layer is eliminated, i.e., the zero elements in the non-zero rows are found. Meanwhile, Fig. 4 (b) reveals the robustness of the proposed algorithm to the in-row sparsity, i.e., the proposed algorithm has stable NMSE performance regardless of whether the non-zero rows are sparse or dense. From Fig. 4 (b), it is clear that as the in-row sparsity increases, the NMSE performance of Proposed TEHE and Proposed A-TEHE remains almost constant and close to which in Oracle-LS even for approaching a spatial stationary channel. In contrast, the Block-SAMP shows a different NMSE performance due to the in-row sparsity.

To further evaluate the performance between the two layers of processing, we define the success recovery rate of the position of non-zero and zero elements to evaluate the performance of the algorithms. First, the sets of the position of non-zero
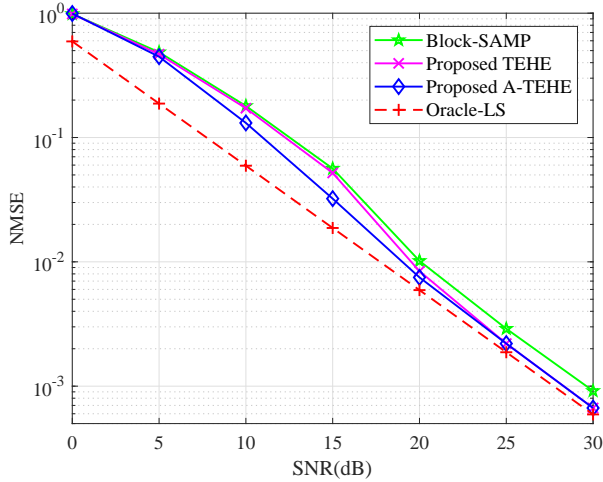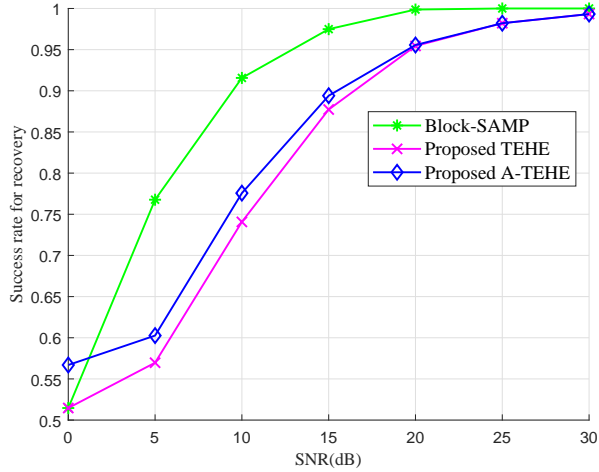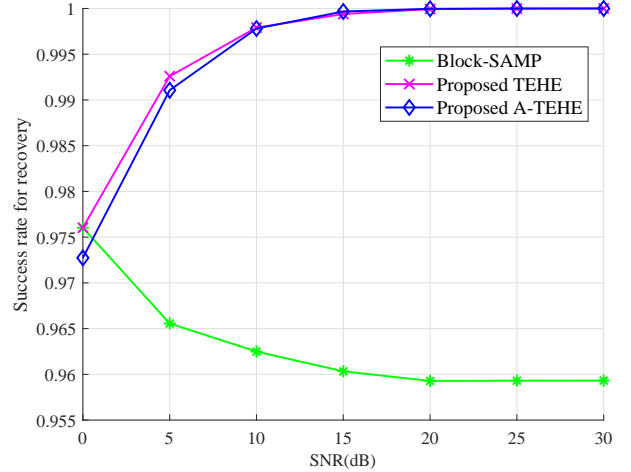
(a) NMSE vs. SNR



(b) NMSE vs. in-row sparsity at SNR=30dB

Fig. 4.  NMSE performance comparison of different method in the second layer.



(a) Position estimation of non-zero elements



(b) Position estimation of zero elements

Fig. 5.  The success rate for recovery of non-zero and zero elements in the second layer.

and zero elements in matrix $\mathbf{X}$ can be written as

$$\mathbb{X}_{nz} = \{(i,j)|x_{i,j} \neq 0\}, \mathbb{X}_z = \{(i,j)|x_{i,j} = 0\}. \quad (38)$$

Then, the success rate for recovery of the position of non-zero and zero elements can be expressed as $\frac{|\hat{\mathbf{X}}(\mathbb{X}_{nz})|_{nz}}{|\mathbf{X}|_{nz}}$ and $\frac{|\hat{\mathbf{X}}(\mathbb{X}_z)|_z}{|\mathbf{X}|_z}$, respectively, where $\hat{\mathbf{X}}(\mathbb{X}_{nz})$ denotes the elements in the matrix $\hat{\mathbf{X}}$ corresponding to the position in $\mathbb{X}_{nz}$, $|\cdot|_{nz}$ and $|\cdot|_z$ denotes the number of non-zero elements and zero elements in the data, respectively. Fig. 5 (a) shows the success recovery rate of non-zero elements. It can be observed that the non-zero success recovery rate increases with increasing SNR and the block-SANP algorithm arrives at 100% with SNR $\geq$ 20dB. This verifies the result in Theorem 2, which states that the first layer algorithm can estimate all correct non-zero rows when the signal is strong enough. In addition, it shows that the non-zero success recovery rate of proposed TEHE and A-TEHE using threshold processing in the second layer is lower than that of block-SAMP. The gap between block-SAMP and the

proposed algorithm becomes smaller at high SNR. From Fig. 5 (b), one can observe that the zero element success recovery rate of block-SAMP cannot increase with the SNR raising, and which can be almost 100% in the proposed TEHE and A-TEHE. This can be explained as that the TEHE and A-TEHE can correctly estimate the positions of zero elements and block-SAMP can not remove the noise in the position of zero elements. Therefore, the NMSE performance of TEHE and A-TEHE outperforms that of block-SAMP.

### C. Impact of the Row Sparsity

In Fig. 6, the effect of the row sparsity of the channel matrix $\mathbf{X}$ and step size is presented. Due to the exact recovery condition, two cases of SNR=15dB and SNR=30dB are discussed. It can be seen that the estimation at SNR=15dB setting step size as one is poor. The reason is that the matrix does not satisfy the exact recovery condition $\|\hat{\mathbf{x}}^{(i)}\|_2 \leq \frac{2\sigma\sqrt{N_p M}}{1-(2k-1)\mu}$ at SNR=15.
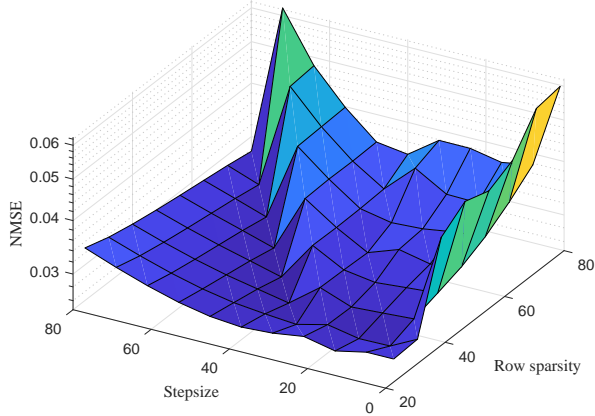
(a) SNR=15dB
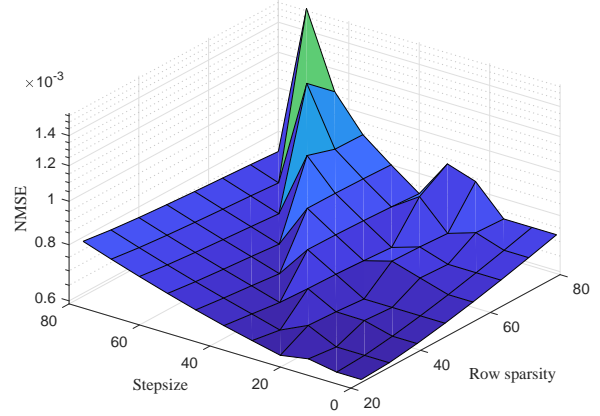


(b) SNR=30dB

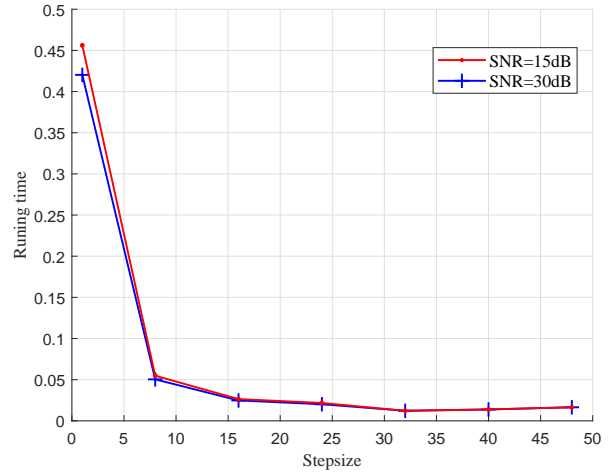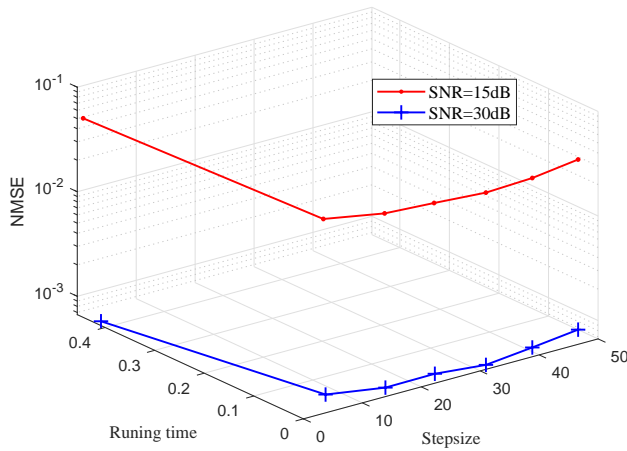Fig. 6. The relationship among step size, row sparsity and NMSE.





Fig. 7. The relationship among step size, running time and NMSE.

Therefore, the correct non-zero rows cannot be found completely, which inevitably leads to unsatisfactory NMSE results. On the contrary, increasing the step size compared to 1 results in extra rows, which compensate for the above disadvantages with a high probability; while SNR=30dB is completely different. Because matrix satisfies the exact recovery condition at SNR=30dB. Except for this one difference, the trends in Fig. 6 (a) and Fig. 6 (b) basically coincide. To be specific, we can divide both graphs Fig. 6 (a) and Fig. 6 (b) into two parts according to the relationship between step size and sparsity to observe. In the case where the step size is less than the row sparsity, i.e, $s < \|\mathbf{X}\|_{2,0}$, at the same sparsity, the larger the step size the worse the NMSE performance, but the difference in NMSE performance between smaller step sizes is negligible. And at the same step length, the greater the row sparsity the worse the NMSE performance. Similarly, in the case where the step size is greater than the row sparsity, i.e, $s > \|\mathbf{X}\|_{2,0}$, the smaller the difference between step size and the row sparsity, the better the NMSE performance. Furthermore, from Fig. 6, we can find that except for some step sizes in the middle of

the figure which are smaller but close to the row sparsity and much larger than the row sparsity, the difference in NMSE performance at other steps is very slight. Fig. 6 illustrates that in the general case we have a relatively large choice of step size.

D. Computational Efficiency

Fig. 7 illustrates the relationship among NMSE performance, running time and step size in the proposed A-TEHE. It is worth noting that the proposed A-TEHE degenerates to the proposed TEHE for the case that the step size is 1. It can be clearly seen from Fig. 7 that the runtime decreases with increasing step size. The step size is inversely proportional to the running time. Moreover, as shown in Fig. 7, the NMSE performance of A-TEHE with $s = 1$ is the worst at SNR=15dB, but it is improved at SNR=30dB. This can be explained by the fact that the channel matrix $\mathbf{X}$ can not satisfy the exact recovery condition for SNR $\leq$15dB. As a result, properly increasing the step size not only improves the speed of the algorithm but also improves the NMSE performance
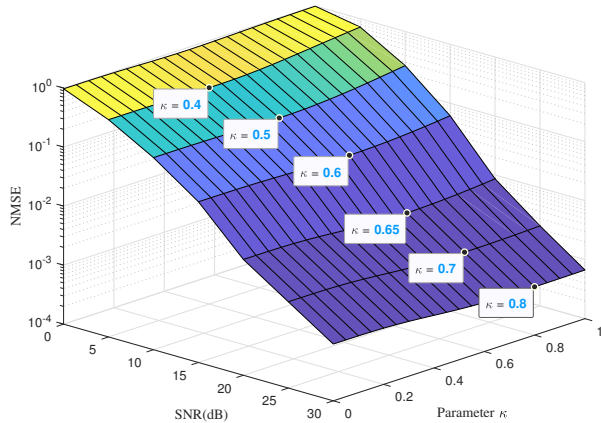
(a) MNSE vs. SNR and Parameter $\kappa$.
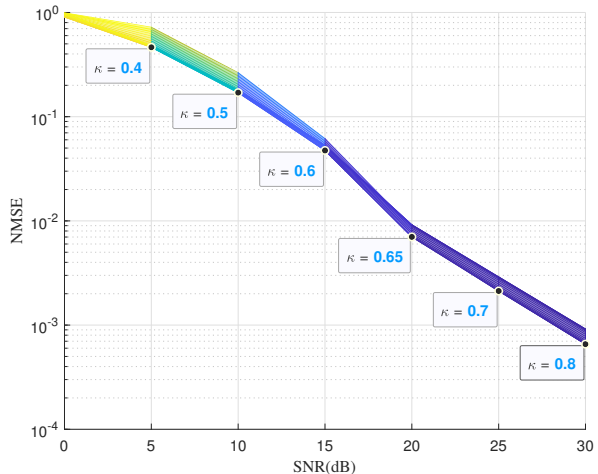


(b) NMSE vs. SNR.

Fig. 8.   The influence of reconstructed noise level parameter $\kappa$.

for low SNR. Naturally, the value of the channel matrix $\mathbf{X}$ satisfies the exact recovery condition at SNR=30dB, i.e., the power of the noise is low enough, at the NMSE performance of A-TEHE with $s = 1$ is the same as the case that the step size is increased appropriately. On the other hand, in Fig. 7, we also can find that NMSE performance weakens slightly with further increases in step size. The increase in step length leads to an increase in the redundant support set, and although the second layer of processing of the algorithm can eliminate some of the effects, the NMSE performance still consists of a small difference.

### E. Impact of Reconstructed Noise Level $\kappa$

Fig. 8 reveals the effect of the selection of reconstructed noise level parameter $\kappa$ in (24) on the NMSE performance of the proposed algorithms. The boundaries of the different colored areas in Fig. 8 (a) represent the NMSE performance of different parameters $\kappa$ at each SNR, and it can be seen that these intersections are not straight lines, i.e., different selected $\kappa$ values will bring differences to the NMSE performance. From Fig. 8, on the one hand, the selection of different parameters $\kappa$ at different SNRs will have different effects on the NMSE performance. On the other hand, because $\kappa \in [0, 1]$, which indirectly affects the NMSE by influencing the threshold $\zeta_0$ in Theorem 1 and $\eta_0$ in Corollary 1, it can only cause small-scale performance fluctuations of NMSE. This can be clearly seen in Fig. 8 (b), where different $\kappa$ at the same SNR results in a slight change in NMSE performance. Furthermore, from Fig. 8, we can confirm our previous judgment on $\kappa$, that the choice of $\kappa$ is related to the SNR. The labeled points in Fig. 8 represent the values of $\kappa$ that result in optimal NMSE performance at this SNR. It can be found that as the SNR increases, the optimal value of $\kappa$ also increases. For SNR=30dB, the optimal value of $\kappa$ is close to 1, and the NMSE corresponding to $\kappa$ values larger than it is almost equal.

## VI. Conclusions

In this paper, we investigated the spatial non-stationary channel estimation for the massive MIMO systems. By exploiting the special non-uniform sparse structure, the channel estimation is formulated as a hierarchical sparse signal recovery problem, and the threshold-enhanced hierarchical algorithms including the TEHE and the A-TEHE have been proposed to achieve high-resolution channels and support set estimations. Furthermore, we provided the performance analysis for the proposed channel estimation algorithms. In particular, the accurate estimation condition and the halting condition for the first layer estimation are derived. Simulation results show that the proposed algorithms can significantly improve the estimation performance in terms of NMSE, success rate for recovery, and computation efficiency.

## Appendix A
### The Proof of Theorem 1

The problem (22) can be translated into $|\mathbb{E}|M$ independent problems, which can be expressed as

$$\hat{x}_{\mathbb{E},cn} = \arg \min_{x_{\mathbb{E},cn} \in \mathbb{C}} \frac{1}{2}|\tilde{y}_{cn} - x_{\mathbb{E},cn}|^2 + \zeta \mathbb{I}_{|x_{\mathbb{E},cn}|}, \quad (39)$$

where $\mathbb{I}_{|x_{\mathbb{E},cn}|} = 1$ for $x_{\mathbb{E},cn} \neq 0$, otherwise, $\mathbb{I}_{|x_{\mathbb{E},cn}|} = 0$. It is important to point out that $\frac{1}{2}|\tilde{y}_{cn} - x_{\mathbb{E},cn}|^2$ and $\zeta \mathbb{I}_{|x_{cn}|}$ are non-negative functions, i.e., $\frac{1}{2}|\tilde{y}_{cn} - x_{\mathbb{E},cn}|^2 \geq 0$ and $\zeta \mathbb{I}_{|x_{cn}|} \geq 0$. Since $\tilde{y}_{cn} \neq 0$, we can not find a optimal solution $\hat{x}_{\mathbb{E},n}$ such that $\frac{1}{2}|\tilde{y}_{cn} - x_{\mathbb{E},n}|^2 = 0$ and $\zeta \mathbb{I}|x_{\mathbb{E},cn}| = 0$ at the same time. Therefore, the minimum value of objective function is obtained from one of $\frac{1}{2}|\tilde{y}_{cn} - x_{\mathbb{E},cn}|^2$ and $\zeta \mathbb{I}_{|x_{\mathbb{E},cn}|}$. Note that the maximum value of $\frac{1}{2}|\tilde{y}_{cn} - x_{\mathbb{E},cn}|^2$ is $|\tilde{y}_{cn}|^2$ when $x_{\mathbb{E},cn} = 0$, and the maximum value of $\zeta \mathbb{I}_{|x_{\mathbb{E},cn}|}$ is $\zeta$ when $x_{\mathbb{E},cn} \neq 0$. The optimal solution of problem (39) can be obtained by considering the following two cases:

- If $\zeta \geq |\tilde{y}_{cn}|^2$, it is worth to set $x_{\mathbb{E},cn} = 0$, and the objective value is $|\tilde{y}_{cn}|_2^2$.

- Otherwise, i.e., $0 < \zeta < |\tilde{y}_{cn}|^2$, $x_{\mathbb{E},cn} = \tilde{y}_{cn}$, and the objective value is $\zeta$.

Summarily, a closed-form solution of problem (39) is given by the following hard shrinkage operator

$$\hat{x}_{\mathbb{E},cn} = \text{shrink}(\tilde{y}_{cn}, \zeta) \triangleq \begin{cases} 0, & \text{if} \quad \zeta \geq |\tilde{y}_{cn}|^2, \\ \tilde{y}_{cn}, & \text{if} \quad 0 < \zeta < |\tilde{y}_{cn}|^2. \end{cases} \quad (40)$$

Recall that the received signal in (20) is only impacted by the Gaussian noise, and the difference between the real channel and the estimated channel can be expressed as

$$\tilde{y}_{cn} - x_{\mathbb{E},cn} = \tilde{w}_{cn}. \quad (41)$$

It means that the estimated signal of the first layer differs from the original signal only by the noise. In other words, the energy of the reconstructed noise and the received signal can be used to design the threshold $\zeta$. Especially, if the energy of the received signal is greater than the upper bound of the energy of the reconstructed noise, i.e., $|\tilde{y}_{cn}|^2 > \sup(|\tilde{w}_{cn}|^2)$, we can make sure $x_n \neq 0$. On the other hand, it can be judged that $x_n = 0$ if the power of the received signal is less than the upper bound of the power of the reconstructed noise, i.e., $|\tilde{y}_{cn}|^2 \leq \sup(|\tilde{w}_{cn}|^2)$. Therefore, the mean of the upper bound of reconstructed Gaussian noise energy can be used to design the threshold $\zeta$. Note that the energy of reconstructed Gaussian noise is

$$\begin{aligned} ||\tilde{\mathbf{W}}||_F^2 &= ||(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})^{-1} \mathbf{S}_{\mathbb{E}}^H \mathbf{W}||_F^2 \\ &= \sum_{i=1}^M ||(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})^{-1} \mathbf{S}_{\mathbb{E}}^H \mathbf{w}_i||_2^2 \\ &\leq \sum_{i=1}^M ||(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})^{-1} \mathbf{S}_{\mathbb{E}}^H||_2^2 ||\mathbf{w}_i||_2^2 \\ &= ||(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})^{-1} \mathbf{S}_{\mathbb{E}}^H||_2^2 ||\mathbf{W}||_F^2, \end{aligned} \quad (42)$$

where $\mathbf{w}_i$ is the $i$-th column of $\mathbf{W}$. Then, the average of the upper bound in (42) is given by

$$\begin{aligned} \bar{\zeta} &= ||(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})^{-1} \mathbf{S}_{\mathbb{E}}^H||_2^2 \frac{||\mathbf{W}||_F^2}{|\mathbb{E}|M} \\ &= ||(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})^{-1} \mathbf{S}_{\mathbb{E}}^H||_2^2 \frac{N_p \sigma^2}{|\mathbb{E}|}. \end{aligned} \quad (43)$$

The above parameter $\bar{\zeta}$ is derived on the premise that the signal energy is much larger than the reconstructed noise energy. It is worth noting that when the reconstructed noise energy is similar to the signal, this parameter $\bar{\zeta}$ (average reconstruction noise upper bound) is slightly larger, that is, the signal may also be smaller than this parameter and be judged to be zero. To reveal the impact of the reconstructed noise energy, we design the threshold as the following based on the reconstructed noise level

$$\zeta_0 = ||(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})^{-1} \mathbf{S}_{\mathbb{E}}^H||_2^2 \frac{N_p \sigma^2}{|\mathbb{E}|} \kappa. \quad (44)$$

where $\kappa \in [0,1]$ is a reconstructed noise level parameter.

## APPENDIX B
## THE PROOF OF COROLLARY 1

Since problem (28) is column-wise decomposable, we can basically solve

$$\hat{\mathbf{x}}_r = \arg\min_{\mathbf{x}_r} \frac{1}{2}||\tilde{\mathbf{y}}_r - \mathbf{x}_r||_2^2 + \eta \mathbb{I}_{||\mathbf{x}_r||_2}. \quad (45)$$

Further, we can analyze each element, and (45) can be transformed into

$$\{\hat{x}_{r,i}\}_{i=1}^M = \arg\min_{\{x_{r,i}\}_{i=1}^M} \sum_{i=1}^M \frac{1}{2}||\tilde{y}_{r,i} - x_{r,i}||_2^2 + \eta \mathbb{I}_{|x_{r,i}|}, \quad (46)$$

where $x_{r,i}, \hat{x}_{r,i}, \tilde{y}_{r,i}$ are the $i$-th element in $\mathbf{x}_r$, respectively. It is clear that the basic problem after decomposition of (45) is similar to (39). It is easy to obtain the result (40), which is similar to the result of Theorem 1.

**Remark 5.** *The element-wise estimation layer of the proposed algorithm is used to estimate the non-zero elements from the non-zero rows by solving a constrained $l_{0,1}$-norm optimization problem. Furthermore, we derive its closed-form solution to obtain the results of **Theorem 1** and **Corollary 1**.*

## APPENDIX C
## THE PROOF OF THEOREM 2

Suppose the algorithm selects the correct support set at the $(t-1)$-th update and the set which has been selected at the current step is $\mathbb{E}_{(t-1)} \subseteq \mathbb{E}$. The residual at the $(t-1)$-th update can be written as

$$\begin{aligned} \mathbf{R}^{(t-1)} &= \mathbf{Y} - \mathbf{S}_{\mathbb{E}_{(t-1)}} \mathbf{S}_{\mathbb{E}_{(t-1)}}^\dagger \mathbf{Y} \\ &= (\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t-1)}} \mathbf{S}_{\mathbb{E}_{(t-1)}}^\dagger)\mathbf{SX} + (\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t-1)}} \mathbf{S}_{\mathbb{E}_{(t-1)}}^\dagger)\mathbf{W} \\ &\triangleq \mathbf{H}^{(t-1)} + \mathbf{N}^{(t-1)}, \end{aligned} \quad (47)$$

where $(\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t-1)}} \mathbf{S}_{\mathbb{E}_{(t-1)}}^\dagger)\mathbf{SX} \triangleq \mathbf{H}^{(t-1)}$ is the residual of the CSI and $(\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t-1)}} \mathbf{S}_{\mathbb{E}_{(t-1)}}^\dagger)\mathbf{W} \triangleq \mathbf{N}^{(t-1)}$ is the residual of the noise. Compared to the previous work [21], [23], the proof of the matrix form leads to the problem of calculating the matrix norm. It is difficult to compare the values of the matrix norm with the boundary conditions. By performing several suitable scaling transformations, we can obtain the exact recovery conditions of the sparse matrix. In the next update, i.e., the $(t)$-th update, (10) can be written as

$$c_r^{(t)} = ||\mathbf{s}_r^H(\mathbf{H}^{(t-1)} + \mathbf{N}^{(t-1)})||_1. \quad (48)$$

Therefore, the following correlation coefficient are available when selecting a support set

$$\begin{cases} c_{r,1}^{(t)} = \max_{\mathbf{s} \in \mathbf{S}_{\mathbb{E}}} \{||\mathbf{s}_r^H \mathbf{H}^{(t-1)}||_1\} \\ c_{r,2}^{(t)} = \max_{\mathbf{s} \in \mathbf{S}_{\mathbb{E}C}} \{||\mathbf{s}_r^H \mathbf{H}^{(t-1)}||_1\} \\ c_{r,3}^{(t)} = \max_{\mathbf{s} \in \mathbf{S}} \{||\mathbf{s}_r^H \mathbf{N}^{(t-1)}||_1\}. \end{cases} \quad (49)$$

To ensure that Algorithm 1 can select the correct row support set, the sufficient condition can be given by

$$\max_{\mathbf{s} \in \mathbf{S}_{\mathbb{E}}} \{||\mathbf{s}_r^H \mathbf{R}^{(t-1)}||_1\} > \max_{\mathbf{s} \in \mathbf{S}_{\mathbb{E}C}} \{||\mathbf{s}_r^H \mathbf{R}^{(t-1)}||_1\}. \quad (50)$$

Based on (47) and (49), we have

$$\max_{\mathbf{s}\in\mathbf{S}_{\mathbb{E}}}\{\|\mathbf{s}_r^H\mathbf{R}^{(t-1)}\|_1\} \geq c_{r,1}^{(t)} - c_{r,3}^{(t)}, \tag{51}$$

$$\max_{\mathbf{s}\in\mathbf{S}_{\mathbb{E}C}}\{\|\mathbf{s}_r^H\mathbf{R}^{(t-1)}\|_1\} \leq c_{r,2}^{(t)} + c_{r,3}^{(t)}. \tag{52}$$

Then, the sufficient condition in (50) can be expressed as

$$c_{r,1}^{(t)} - c_{r,2}^{(t)} > 2c_{r,3}^{(t)}. \tag{53}$$

From the Cauchy inequality, we get

$$
\begin{aligned}
c_{r,3}^{(t)} &= \max_{\mathbf{s}_r\in\mathbf{S}}\{\|\mathbf{s}_r^H\mathbf{N}^{(t)}\|_1\} \\
&= \max_{\mathbf{s}_r\in\mathbf{S}}\sum_{i=1}^{M}\{\mathbf{s}_r^H\mathbf{n}_i^{(t)}\} \\
&\leq \max_{\mathbf{s}_r\in\mathbf{S}}\sum_{i=1}^{M}\{\|\mathbf{s}_r^H\|_2\|\mathbf{n}_i^{(t)}\|_2\} \\
&= \max_{\mathbf{s}_r\in\mathbf{S}}\{\|\mathbf{s}_r^H\|_2\|\mathbf{N}^{(t)}\|_F\} \\
&\leq \|\mathbf{s}_r^H\|_2\|\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger}\|_2\|\mathbf{W}\|_F \\
&\overset{(a)}{=} \sqrt{N_pM\sigma^2}, \tag{54}
\end{aligned}
$$

where $\mathbf{n}_i^{(t)}$ is the $i$-th column of $\mathbf{N}^{(t)}$ and $(a)$ is obtained by the fact that $\mathbf{s}_r$ is normalized. In addition, $\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger}$ is idempotent matrix, i.e., $(\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger})^2 = \mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger}$, thence the eigenvalues of this matrix can only be 0 or 1. Obviously, $\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger} \neq 0$, then $\|\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger}\|_2 = 1$. Therefore, the sufficient condition in (50) can be further written as

$$c_{r,1}^{(t)} - c_{r,2}^{(t)} > 2\sigma\sqrt{N_pM}. \tag{55}$$

To specify the relationship between $c_{r,1}^{(t)}$ and $c_{r,2}^{(t)}$, we can consider the noiseless situation i.e., the noise matrix $\mathbf{W}=0$ and $c_{r,3}^{(t)}=0$, define

$$\rho = \frac{c_{r,2}^{(t)}}{c_{r,1}^{(t)}} = \frac{\max_{\mathbf{s}\in\mathbf{S}_{\mathbb{E}C}}\{\|\mathbf{s}_r^H\mathbf{H}^{(t-1)}\|_1\}}{\max_{\mathbf{s}\in\mathbf{S}_{\mathbb{E}}}\{\|\mathbf{s}_r^H\mathbf{H}^{(t-1)}\|_1\}}. \tag{56}$$

We can calculate that

$$
\begin{aligned}
\rho &= \frac{\max_{\mathbf{s}_r\in\mathbf{S}_{\mathbb{E}C}}\{\|(\mathbf{s}_r^H(\mathbf{S}_{\mathbb{E}}^{\dagger})^H\mathbf{S}_{\mathbb{E}}^H\mathbf{H}^{(t-1)})^H\|_1\}}{\max_{\mathbf{s}\in\mathbf{S}_{\mathbb{E}}}\{\|(\mathbf{s}_r^H\mathbf{H}^{(t-1)})^H\|_1\}} \\
&\leq \frac{\max_{\mathbf{s}_r\in\mathbf{S}_{\mathbb{E}C}}\{\|(\mathbf{S}_{\mathbb{E}}^{\dagger})\mathbf{s}_r\|_1\}\max_{\mathbf{s}\in\mathbf{S}_{\mathbb{E}}}\{\|(\mathbf{H}^{(t-1)})^H\mathbf{s}_r\|_1\}}{\max_{\mathbf{s}\in\mathbf{S}_{\mathbb{E}}}\{\|(\mathbf{H}^{(t-1)})^H\mathbf{s}_r\|_1\}} \\
&\leq \max_{\mathbf{s}_r\in\mathbf{S}_{\mathbb{E}C}}\{\|(\mathbf{S}_{\mathbb{E}}^{\dagger})\mathbf{s}_r\|_1\} = 1 - \mathrm{ERC}(\mathbb{E}). \tag{57}
\end{aligned}
$$

Based (56) and (57), we have $c_{r,1}^{(t)} - c_{r,2}^{(t)} \geq \mathrm{ERC}(\mathbb{E})c_{r,1}^{(t)}$. According to Lemma 1, we further have

$$c_{r,1}^{(t)} - c_{r,2}^{(t)} \geq \frac{1-(2k-1)\mu}{1-(k-1)\mu}c_{r,1}^{(t)}, \tag{58}$$

i.e.,

$$c_{r,2}^{(t)} \leq \left(1 - \frac{1-(2k-1)\mu}{1-(k-1)\mu}\right)c_{r,1}^{(t)}. \tag{59}$$

Substitute (59) into (55), the sufficient condition can be given by

$$c_{r,1}^{(t)} > \frac{2-2(k-1)\mu}{1-(2k-1)\mu}\sigma\sqrt{N_pM}. \tag{60}$$

Note that

$$
\begin{aligned}
c_{r,1}^{(t)} &= \max_{\mathbf{s}\in\mathbf{S}_{\mathbb{E}}}\{\|\mathbf{s}_r^H\mathbf{H}^{(t-1)}\|_1\} \geq \max_{\mathbf{s}\in\mathbf{S}_{\mathbb{E}}}\{\|\mathbf{s}_r^H\mathbf{H}^{(t-1)}\|_2\} \\
&= \max_{\mathbf{s}_r\in\mathbf{S}_{\mathbb{E}}}\{\|\mathbf{s}_r^H(\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t-1)}}\mathbf{S}_{\mathbb{E}_{(t-1)}}^{\dagger})\mathbf{S}_{\mathbb{E}}\mathbf{X}^{(\mathbb{E})}\|_2\} \\
&= \max_{\mathbf{s}_r\in\mathbf{S}_{\mathbb{E}}}\{\|\mathbf{s}_r^H(\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t-1)}}\mathbf{S}_{\mathbb{E}_{(t-1)}}^{\dagger})(\mathbf{S}_{\mathbb{E}_{(t-1)}}\mathbf{X}^{\mathbb{E}_{(t)}} \\
&\quad + \mathbf{S}_{\mathbb{E}'_{(t-1)}}\mathbf{X}^{\mathbb{E}'_{(t-1)}})\|_2\} \\
&= \max_{\mathbf{s}_r\in\mathbf{S}_{\mathbb{E}'_{(t-1)}}}\{\|\mathbf{s}_r^H(\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t-1)}}\mathbf{S}_{\mathbb{E}_{(t-1)}}^{\dagger})\mathbf{S}_{\mathbb{E}'_{(t-1)}}\mathbf{X}^{\mathbb{E}'_{(t-1)}}\|_2\} \\
&\triangleq \max_{\mathbf{s}_r\in\mathbf{S}_{\mathbb{E}'_{(t-1)}}}\{\|\mathbf{p}\mathbf{X}^{\mathbb{E}'_{(t-1)}}\|_2\}, \tag{61}
\end{aligned}
$$

where $\mathbf{p} = \mathbf{s}_r^H(\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t-1)}}\mathbf{S}_{\mathbb{E}_{(t-1)}}^{\dagger})\mathbf{S}_{\mathbb{E}'_{(t-1)}}$, $\mathbb{E}'_{(t-1)} = \mathbb{E} - \mathbb{E}_{(t-1)}$, and $\mathbf{X}^{(\mathbb{E}'_{(t)})}$ is the rows corresponding $\mathbb{E}'_{(t-1)}$ in matrix $\mathbf{X}$.

Let $\mathbf{P} = \mathbf{S}_{\mathbb{E}'_{(t)}}^H(\mathbf{I}-\mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger})\mathbf{S}_{\mathbb{E}}$, then we have

$$
\begin{aligned}
c_{r,1}^{(t)} &\geq \frac{\|\mathbf{P}\mathbf{X}^{(\mathbb{E}'_{(t-1)})}\|_F}{\sqrt{k-|\mathbb{E}_{(t-1)}|}} \\
&\geq \frac{\lambda_{\min}(\mathbf{P})\|\mathbf{X}^{(\mathbb{E}'_{(t-1)})}\|_F}{\sqrt{k-|\mathbb{E}_{(t-1)}|}} \\
&\overset{(a)}{\geq} \frac{\lambda_{\min}(\mathbf{S}_{\mathbb{E}}^H\mathbf{S}_{\mathbb{E}})\|\mathbf{X}^{(\mathbb{E}'_{(t-1)})}\|_F}{\sqrt{k-|\mathbb{E}_{(t-1)}|}} \\
&\overset{(b)}{\geq} \frac{(1-(k-1)\mu)\|\mathbf{X}^{(\mathbb{E}'_{(t-1)})}\|_F}{\sqrt{k-|\mathbb{E}_{(t-1)}|}}, \tag{62}
\end{aligned}
$$

where (a) and (b) are obtained from (35) and (56), respectively. According to (60) and (62), the sufficient condition can be expressed as

$$\|\mathbf{X}^{(\mathbb{E}'_{(t)})}\|_F > \frac{2\sqrt{k-|\mathbb{E}_{(t-1)}|}\sigma\sqrt{N_pM}}{1-(2k-1)\mu}. \tag{63}$$

Therefore, we can obtain a sufficient condition related to the noise power, the mutual incoherence $\mu$ and the sparsity $k$ as following

$$\|\hat{\mathbf{x}}^{(i)}\|_2 > \frac{2\sigma\sqrt{N_pM}}{1-(2k-1)\mu}, \tag{64}$$

where $\hat{\mathbf{x}}^{(i)}$ is the $i$-th non-zero row in the matrix $\mathbf{X}$.

## APPENDIX D
### THE PROOF OF THEOREM 3

For our algorithm, we have to study the halting condition. In the previous work [26], [30], the halting condition is mostly empirical, while we analyze and calculate it quantitatively to obtain a solution. Note that the residual between the received signal and the projection of the reconstructed signal on the

matrix $\mathbf{S}_{\hat{\mathbb{E}}}$, i.e. $\mathbf{R} = \mathbf{Y} - \mathbf{S}_{\hat{\mathbb{E}}}\tilde{\mathbf{Y}}$ in the proposed algorithm can be used for the measure of the accuracy of row support estimation. If the support set is found exactly, i.e., $\mathbf{S}_{\hat{\mathbb{E}}} = \mathbf{S}_{\mathbb{E}}$, it is clear that

$$
\begin{aligned}
\mathbf{R} &= (\mathbf{I} - \mathbf{S}_{\mathbb{E}}\mathbf{S}_{\mathbb{E}}^{\dagger})\mathbf{Y} \\
&= (\mathbf{I} - \mathbf{S}_{\mathbb{E}}\mathbf{S}_{\mathbb{E}}^{\dagger})(\mathbf{S}_{\mathbb{E}}\mathbf{X}^{(\mathbb{E})} + \mathbf{W}) \\
&= (\mathbf{I} - \mathbf{S}_{\mathbb{E}}\mathbf{S}_{\mathbb{E}}^{\dagger})\mathbf{W}.
\end{aligned} \tag{65}
$$

It is easy to obtain that the upper bound of the power of the residuals in the this is

$$
\|\mathbf{R}\|_F^2 \leq \|\mathbf{I} - \mathbf{S}_{\mathbb{E}}\mathbf{S}_{\mathbb{E}}^{\dagger}\|_2^2 \|\mathbf{W}\|_F^2 \leq N_p M \sigma^2. \tag{66}
$$

For the case that the support set is not completely found, i.e., in the $(t)$-th update, only part of the correct support sets were found and let $\mathbb{E}_{(t)}$ denote the above set. In this situation, the power residual can be written as

$$
\begin{aligned}
\|\mathbf{R}\|_F^2 &= \|(\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger})(\mathbf{S}_{\mathbb{E}}\mathbf{X}^{(\mathbb{E})} + \mathbf{W})\|_F^2 \\
&\geq \|(\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger})\mathbf{S}_{\mathbb{E}'_{(t)}}\mathbf{X}^{(\mathbb{E}'_{(t)})}\|_F^2 \\
&\quad - \|(\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger})\mathbf{W}\|_F^2 \\
&\geq \|(\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger})\mathbf{S}_{\mathbb{E}'_{(t)}}\mathbf{X}^{(\mathbb{E}'_{(t)})}\|_F^2 - N_p M \sigma^2. \quad (67)
\end{aligned}
$$

It follows from (35) and (56) that

$$
\begin{aligned}
&\|(\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger})\mathbf{S}_{\mathbb{E}'_{(t)}}\mathbf{X}^{(\mathbb{E}'_{(t)})}\|_F^2 \\
&\geq \lambda_{min}(\mathbf{S}_{\mathbb{E}}^H \mathbf{S}_{\mathbb{E}})\|\mathbf{X}^{(\mathbb{E}'_{(t)})}\|_F^2 \\
&\geq (1 - (|\hat{\mathbf{x}}^{(t)}|_0 - 1)\mu)\frac{2\sigma^2 N_p M}{1 - (2|\hat{\mathbf{x}}^{(t)}|_0 - 1)\mu} \\
&> 2\sigma^2 N_p M.
\end{aligned} \tag{68}
$$

From (67) and (68), we can learn that

$$
\begin{aligned}
\|\mathbf{R}\|_F^2 &\geq \|(\mathbf{I} - \mathbf{S}_{\mathbb{E}_{(t)}}\mathbf{S}_{\mathbb{E}_{(t)}}^{\dagger})\mathbf{S}_{\mathbb{E}'_{(t)}}\mathbf{X}^{(\mathbb{E}'_{(t)})}\|_F^2 - N_p M \sigma^2 \\
&> \sigma^2 N_p M.
\end{aligned} \tag{69}
$$

The obtained results in (66) and (69) illustrate that $\|\mathbf{R}\|_F^2 \leq N_p M \sigma^2$ can be used as a halting condition.

## REFERENCES

[1] C. Tan, D. Cai, F. Fang, J. Shan, Y. Xu, Z. Ding, and P. Fan, "Hierarchical sparse estimation of non-stationary channel for uplink massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2023, to be published.

[2] B. M. Lee and H. Yang, "Energy-efficient massive MIMO in massive industrial internet of things networks," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3657–3671, 2022.

[3] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, 2016.

[4] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys Tuts*, vol. 21, no. 4, pp. 3109–3132, 2019.

[5] Z. Ding and H. V. Poor, "On the application of BAC-NOMA to 6G umMTC," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2678–2682, 2021.

[6] E. Björnson and L. Sanguinetti, "Power scaling laws and near-field behaviors of massive MIMO and intelligent reflecting surfaces," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1306–1324, 2020.

[7] A. Zaib, M. Masood, A. Ali, W. Xu, and T. Y. Al-Naffouri, "Distributed channel estimation and pilot contamination analysis for massive MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4607–4621, 2016.

[8] J. Chen, S. Chen, Y. Qi, and S. Fu, "Intelligent massive MIMO antenna selection using monte carlo tree search," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5380–5390, 2019.

[9] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[10] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[11] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, 2020.

[12] Y. Xu, C. Shen, Z. Ding, X. Sun, S. Yan, G. Zhu, and Z. Zhong, "Joint beamforming and power-splitting control in downlink cooperative SWIPT NOMA systems," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4874–4886, 2017.

[13] E. D. Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, 2020.

[14] Q. Zou, H. Zhang, D. Cai, and H. Yang, "Message passing based joint channel and user activity estimation for uplink grant-free massive MIMO systems with low-precision ADCs," *IEEE Signal Process. Lett.*, vol. 27, pp. 506–510, 2020.

[15] W. Shen, L. Dai, J. An, P. Fan, and R. W. Heath, "Channel estimation for orthogonal time frequency space (OTFS) massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4204–4217, 2019.

[16] X. Cheng, K. Xu, J. Sun, and S. Li, "Adaptive grouping sparse bayesian learning for channel estimation in non-stationary uplink massive MIMO systems," *IEEE Trans. Wireless Commun*, vol. 18, no. 8, pp. 4184–4198, 2019.

[17] A. Ali, E. D. Carvalho, and R. W. Heath, "Linear receivers in non-stationary massive MIMO channels with visibility regions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 885–888, 2019.

[18] H. Jiang, C. Ruan, Z. Zhang, J. Dang, L. Wu, M. Mukherjee, and D. B. d. Costa, "A general wideband non-stationary stochastic channel model for intelligent reflecting surface-assisted MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5314–5328, 2021.

[19] Y. Zhu, H. Guo, and V. K. N. Lau, "Bayesian channel estimation in multi-user massive MIMO with extremely large antenna array," *IEEE Trans. Signal Process.*, vol. 69, pp. 5463–5478, 2021.

[20] J. Dai, A. Liu, and H. C. So, "Non-uniform burst-sparsity learning for massive MIMO channel estimation," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1075–1087, 2019.

[21] J. A. Tropp, *Topics in sparse approximation*. The University of Texas at Austin, 2004.

[22] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[23] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[24] D. Cai, P. Fan, Q. Zou, Y. Xu, Z. Ding, and Z. Liu, "Active device detection and performance analysis of massive non-orthogonal transmissions in cellular internet of things," *SCIENCE CHINA Inf. Sci.*, vol. 1, no. 1, pp. 1–17, 2020.

[25] Z. Wan, Z. Gao, B. Shim, K. Yang, G. Mao, and M.-S. Alouini, "Compressive sensing based channel estimation for millimeter-wave full-dimensional MIMO with lens-array," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 2337–2342, 2020.

[26] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, 2016.

[27] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, 2016.

[28] Y. Han, S. Jin, C.-K. Wen, and X. Ma, "Channel estimation for extremely large-scale massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 633–637, 2020.

[29] A. Manoj and A. P. Kannu, "Multi-user millimeter wave channel estimation using generalized block OMP algorithm," *in Proc. IEEE Workshop Signal Process. Adv. Wireless Commun., (SPAWC)*, pp. 1–5, 2017.

[30] S. Hou, Y. Wang, T. Zeng, and S. Wu, "Sparse channel estimation for spatial non-stationary massive MIMO channels," *IEEE Commun. Lett.*, vol. 24, no. 3, pp. 681–684, 2020.

[31] M. Qiu, K. Cao, D. Cai, Z. Dong, and Y. Cui, "Low-complexity joint estimation for asynchronous massive internet of things: An ADMM approach," in *2022 2nd International Conference on Intelligent Technology and Embedded Systems (ICITES)*, 2022, pp. 13–20.

[32] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proc. Asilomar Conf. Signals, Syst. and Comput.*, 2008, pp. 581–587.

[33] L. Liu and W. Yu, "Massive connectivity with massive MIMO part-I: Device activity detection and channel estimation," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2933–2946, 2018.

[34] ——, "Massive connectivity with massive MIMO part-II: Achievable rate characterization," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2947–2959, 2018.

**Chongyang Tan** (S'22) received the B.S. degree from the Henan University of Technology, Zhengzhou, China, in 2021. He is currently pursuing the M.S. degree in cyberspace security at the College of Information Science and Technology of Jinan University. His research interests include signal processing algorithms designs for massive MIMO systems, deep learning for massive MIMO systems and physical layer security.

**Donghong Cai** (M'20) received the B.S. degree from the School of Mathematics and Information Sciences, Shaoguan University, Shaoguan, China, in 2012, and the M.S. and Ph.D. degrees from Southwest Jiaotong University, Chengdu, China, in 2015 and 2020, respectively. From October 2017 to October 2018, he was a visiting Ph.D. student with Lancaster University, Lancaster, U.K., and the University of Manchester, Manchester, U.K. He served as a Guest Editor for a special issues in Physical Communication. He is currently a Lecturer with the College of Information Science and Technology/College of Cyber Security, Jinan University, Guangzhou, China. His current research interests include signal detection, security coding, distributed Internet of Things, privacy preservation, machine learning, and nonorthogonal multiple access.
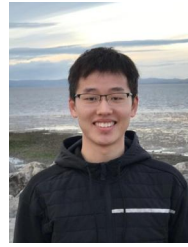
**Yanqing Xu** (M'19) received the Ph.D. degree in communication and information system from the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China, in 2019. He was a Visiting Student with The Chinese University of Hong Kong, Shenzhen, China, from March 2017 to September 2017, and Lancaster University, Lancaster, U.K., from October 2017 to October 2018. He was a senior engineer with Huawei Technologies Company Ltd., from July 2019 to July 2020. From September 2020 to August 2022, he was a PostDoc researcher with The Chinese University of Hong Kong, where he is currently working as a research associate. He is a recipient of Top 3% Paper Recognition of the IEEE ICASSP 2023, the Shenzhen Overseas High-Caliber Personnel (Level C) 2021 and the Huawei Technical Cooperation Achievement Transformation Award (2nd Prize) in 2022. His current research interests are distributed signal processing algorithms designs for massive MIMO systems and federated learning.
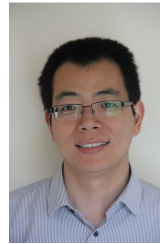
**Zhiguo Ding** (S'03-M'05-F'20) received his B.Eng from the Beijing University of Posts and Telecommunications in 2000, and the Ph.D degree from Imperial College London in 2005. He is currently a Professor in Communications at Khalifa University, and has also been affiliated with the University of Manchester and Princeton University.

Dr Ding's research interests are B5G networks, game theory, cooperative and energy harvesting networks and statistical signal processing. He is serving as an Area Editor for *IEEE Transactions on Wireless Communications*, and *IEEE Open Journal of the Communications Society*, an Editor for *IEEE Transactions on Vehicular Technology* and *IEEE Communications Surveys & Tutorials*, and was an Editor for *IEEE Wireless Communication Letters*, *IEEE Transactions on Communications*, *IEEE Communication Letter*s from 2013 to 2016. He recently received the EU Marie Curie Fellowship 2012-2014, the Top IEEE TVT Editor 2017, IEEE Heinrich Hertz Award 2018, IEEE Jack Neubauer Memorial Award 2018, IEEE Best Signal Processing Letter Award 2018, Friedrich Wilhelm Bessel Research Award 2020, and IEEE SPCC Technical Recognition Award 2021. He is a Fellow of the IEEE, a Distinguished Lecturer of IEEE ComSoc, and a Web of Science Highly Cited Researcher in two categories 2022.

**Pingzhi Fan** (M'93-SM'99-F'15) received the M.Sc. degree in computer science from Southwest Jiaotong University, China, in 1987, and the Ph.D. degree in electronic engineering from Hull University, U.K., in 1994. He is currently presidential professor of Southwest Jiaotong University (SWJTU), honorary dean of the SWJTU-Leeds Joint School, and a visiting professor of Leeds University, UK (1997-). He served as an EXCOM member for the IEEE Region 10, IET (IEE) Council, and the IET Asia Pacific Region. He is a recipient of the UK ORS Award (1992), the National Science Fund for Distinguished Young Scholars (1998, NSFC), IEEE VT Society Jack Neubauer Memorial Award (2018), IEEE SP Society SPL Best Paper Award (2018), IEEE/CIC ICCC2020 Best Paper Award, IEEE WCSP2022 Best Paper Award, and IEEE ICC2023 Best Paper Award. He served as a chief scientist of the National 973 Plan Project (National Basic Research Program of China) between 2012.1-2016.12. His research interests include high mobility wireless communications, massive random-access techniques, signal design & coding, etc. He is an IEEE VTS Distinguished Speaker (2019-2025), a fellow of IEEE, IET, CIE and CIC.