

Received October 24, 2021, accepted November 11, 2021, date of publication November 17, 2021, date of current version January 14, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3128929

# Tensor Dictionary Manifold Learning for Channel Estimation and Interference Elimination of Multi-User Millimeter-Wave Massive MIMO Systems

XIAOPING ZHOU<sup>ID</sup>, HAICHAO LIU<sup>ID</sup>, BIN WANG<sup>ID</sup>, JIFENG HUANG, AND YANG WANG<sup>ID</sup>

Shanghai Normal University, Shanghai 200234, China

Corresponding authors: Xiaoping Zhou (zxpshnu@163.com) and Bin Wang (wangbin527.2008@163.com)

This work was supported by the Shanghai Capacity Building Projects in Local Institutions under Grant 19070502900.

**ABSTRACT** Millimeter Wave (mmWave) massive multiple-input multiple-output (MIMO) systems with hybrid analog-digital architectures can greatly increase system capacity and communicate with multiple users at the same time. Accurate channel estimation is crucial for multi-user communications, but its accuracy is limited as the number of antennas and users increases. In the matrix high-dimensional operation for multi-user channel estimation, not only is it computationally intensive, but also the estimation accuracy is low. It makes our work turn to channel estimation of the user group within a certain region to improve the accuracy of estimation. In this paper, we propose a tensor dictionary manifold learning method for channel estimation and interference elimination of the multi-user mmWave massive MIMO system. A multi-user digital-analog mixed received signal model is presented. The tensor dictionary manifold learning scheme is proposed to model the received signal as a third-order low-rank tensor to handle the high-dimensional user, antenna, and channel. After segmentation, clustering and manifold learning, multiple tensor dictionary manifold models containing a group of user signals are fitted. Tensor dictionary manifold learning can take advantage of the inherent multi-domain properties of signals in the frequency, time, code and spatial domains to maintain inter-user correlation within a user group while reducing the high-dimensional channels of the user group. Using the convex relaxation property of the tensor alternating direction method, we propose a strategy to eliminate interference from other groups. And with the help of the multi-signal classification method, the channel parameters of user groups are obtained to improve the accuracy of multi-user channel estimation. This method can perform channel estimation for multiple users with only a few pilots, and improve the performance of the system. Numerical results confirm the good performance of this method.

**INDEX TERMS** MIMO, channel estimation, tensor, dictionary learning, manifold, ADMM, MUSIC.

## I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) is the key technology of the future 5G wireless communication system [1]. It configures a large number of antenna elements in a small space to obtain large multiplexing gain [2] and improve the capacity of wireless communication [3]. Therefore, the massive MIMO system supports communication with multiple users. However, the communication between multiple users needs to ensure strict service quality and

requires advanced signal processing techniques [4], [5], such as channel estimation, channel equalization and channel coding. Channel estimation is particularly important for massive MIMO systems. With the increase of antennas, accurate and effective channel estimation is a research hot and challenging problem [6], [7].

Channel estimation is one of the most critical issues in wireless communication, and its ultimate goal is to estimate the wireless channel as accurately as possible under a limited number of pilot signals. Under different channel models and wireless systems, such as multi-user massive MIMO [8], millimeter-wave (mmWave) massive MIMO [9] and the

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

Internet of Things (IoT) [10], the challenges are significantly different. Therefore, people use different channel estimation methods to apply in different scenarios. By exploiting the hidden joint sparsity in the multi-user channel matrix, [11] proposed a joint orthogonal matching tracking recovery algorithm, in which the channel matrix shares some common scatterers and common sparsity among geographically adjacent users. Multiple signal classification (MUSIC) [12] and rotation invariance technology (ESPRIT) [13] are used to process array signals to improve the performance of channel estimation [14]–[16]. These search signal methods are based on the location information of the user source. [17] proposes an Interior Point (IP) aided orthogonal matching pursuit (OMP) algorithm. It significantly improves the channel estimation accuracy by reducing the estimation error of angle of departure (AoD) and angle of arrival (AoA). A performance measure called the squared position error bound is proposed to characterize localization accuracy [18]. The authors consider the multicarrier signals for source localization in [19]. The time angle of departure (TAoD) estimates is obtained from the cyclic prefixes inserted before each orthogonal frequency division multiplexing (OFDM) data block. These estimates are then used for localization. Indoor radio positioning based on AoD has also been studied in [20] which applies the MUSIC algorithm on the signals from noncoherent antenna arrays, and in [19] the authors consider AoD estimation from channel state information in MIMO-OFDM systems. However, previous studies on massive MIMO communication [21]–[24] mainly focused on single-user channel estimation. The scenario where the base station (BS) estimates multiple user channels at the same time is not considered.

The signal processing technology based on tensor can make use of the inherent multi-domain characteristics of the signal in the frequency, time, code and space domains. It obtains the optimal solution through tensor fitting on the basis of decomposition uniqueness [25]–[27]. At the same time, tensor decomposition will not destroy the internal relationship of each element, and can make full use of the spatial structure information of the signal, thereby improving the estimation accuracy. In addition, tensor modeling [28]–[30] benefits from multiple diversity. This feature is helpful to achieve multi-user signal separation, equalization and channel estimation under more relaxed identifiability conditions than the traditional matrix method. It is interesting to note that, low-rank tensor decomposition-aided channel estimation for mmWave MIMO-OFDM systems is developed in [31]. However, for each dimension, one dimensional search is required and the complexity of the spectral search may still be unacceptably high for real-time problems. In [32], a multi-linear singular value decomposition (MSVD) method is used for channel parameter estimation. The Tucker form of the MSVD enables paths to be extracted based on signal energy. But this method cannot be used in a multi-user scenario. [33] derive a spatial-frequency channel model that incorporates the multipath parameters. And spatial smoothing method and structured CP decomposition are used for channel estimation.

The method does not benefit much from the phase rarefaction effect due to its frequency-independent property. In [34], the multi-user channel estimation problem was investigated through the lens of Bayesian tensor methods. [35] use a generalization of beamspace-ESPRIT method from matrix to tensor framework. The channel parameters are automatically associated. But the effect of the association is not satisfactory.

In this paper, we consider the tensor dictionary manifold learning for channel estimation and interference elimination of the multi-user mmWave massive MIMO system. With the multi-user mmWave massive MIMO system with a mixed digital-analog structure greatly increasing the system capacity, problems such as inaccurate channel estimation and low resolution for multi-users have emerged. Accurate channel estimation is crucial in multi-user communication, which drives our research and makes us turn to channel estimation of the user group to improve resolution. For multi-user scenarios, we developed a tensor manifold dictionary learning model. The received signals of all users are divided as a third-order low-rank tensor and clustered to form a dictionary learning (DL) model of the received signals of multiple user groups. Through manifold learning, the relationship between adjacent users in the group is analyzed, and the high-dimensional channels in the group are embedded in the low-dimensional space. Using the convex relaxation properties of tensor Alternating direction method of multiplier (ADMM), the interference of other user groups after clustering can be eliminated. And with the help of the MUSIC method, the channel parameters of the user group are obtained. A high-precision channel estimation of the user group is realized. This method can perform channel estimation for multiple users with only a few pilots, and improve the performance of the system. Numerical results confirm the good performance of this method.

The remainder of this paper is organized as follows. Section II introduces the notions utilized and preliminary work in the paper. In Section III, the multi-user massive MIMO system and channel model are presented. Section IV describes segmentation and clustering of received signals for dictionary learning. Section V presents the manifold learning method for analyzing the relationship. Section VI gives the tensor ADMM method and the MUSIC method to get the channel estimation of the user groups. Some numerical results are provided in Section VII. Finally, we conclude this paper in Section VIII.

## II. NOTIONS AND PRELIMINARY WORK

In this section, we introduce the notions utilized and the preliminary work related to tensors in the paper. More detailed instructions are in [36]–[38].

### A. NOTIONS

Bold lower-case and upper-case letters are used for vectors and matrices, respectively, while regular letters denote scalars. Tensor is represented by curlicue letters ( $\mathcal{A}, \mathcal{B}, \dots$ ).  $(\cdot)^{-1}$ ,  $(\cdot)^T$ ,  $(\cdot)^H$  and  $(\cdot)^*$  represent inverse, transpose,

conjugate transpose and complex conjugate, respectively.  $\circ$ ,  $*$ ,  $\otimes$  and  $\odot$  mean out product, Hadamard product, Kronecker product and Khatri-Rao product, respectively.  $\|\cdot\|_F$ ,  $E[\cdot]$  represent Frobenius norm and expectation.

## B. PRELIMINARY WORK

A tensor of order  $N$ , corresponding to an  $N$ -dimensional data matrix, is denoted as  $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_n \times \dots \times I_N}$ . Elements of  $\mathcal{A}$  are denoted as  $a_{i_1 \dots i_n \dots i_N}$ , where  $1 \leq i_n \leq I_n$ . The mode- $n$  vectors of an  $N$  order tensor  $\mathcal{A}$  are the  $I_n$  dimensional vectors obtained from  $\mathcal{A}$  by varying index  $i_n$  while keeping the other indices fixed. The matrix  $\mathcal{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots I_N)}$  is composed by taking the mode- $n$  vectors of  $\mathcal{A}$  as its columns. This matrix can also be naturally seen as the mode- $n$  flattening of the tensor  $\mathcal{A}$ . The n-rank of  $\mathcal{A}$ , denoted as  $r_n$ , is the dimension of the vector space spanned by the mode- $n$  vectors of  $\mathcal{A}$ .

The product of two matrices can be generalized to the product of a tensor and a matrix. The mode- $n$  product of a tensor  $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_n \times \dots \times I_N}$  by a matrix  $\mathcal{B} \in \mathbb{R}^{J_n \times I_n}$ , denoted by  $\mathcal{A} \times_n \mathcal{B}$ , is also an  $N$  order tensor  $\mathcal{C} \in \mathbb{R}^{I_1 \times \dots \times J_n \times \dots \times I_N}$ . The mode- $n$  product  $\mathcal{C} = \mathcal{A} \times_n \mathcal{B}$  can also be calculated by the matrix multiplication  $\mathcal{C}_{(n)} = \mathcal{B}\mathcal{A}_{(n)}$ , followed by a re-tensorization of undoing the mode- $n$  flattening. The Frobenius norm of a tensor  $\mathcal{A}$  is defined as:

$$\|\mathcal{A}\|_F = \left( \sum_{i_1, \dots, i_N} |a_{i_1 \dots i_N}|^2 \right)^{1/2}.$$

## III. SYSTEM AND CHANNEL MODEL

In this section, we introduce the multi-user massive mmWave MIMO system and channel model.

### A. SYSTEM MODEL

We consider a multi-user mmWave massive MIMO system consisting of a base station (BS) and  $K$  multiple mobile stations (MSs), i.e., users. It is depicted in Fig. 1. To facilitate the hardware implementation, hybrid analog and digital beamforming structures (Fig. 1) are employed by both the BS and the MS. The antenna array adopts uniform linear array (ULA). We assume that the BS is equipped with  $N_t$  antennas

and  $N_{RF}^t$  RF chains, and the MS is equipped with  $N_r$  antennas and  $N_{RF}^r$  RF chains.  $N_t$  and  $N_r$  antennas are divided into  $N_\delta$  groups. Therefore, each MS supports  $N_{RF}^r$  data streams. Without loss of generality, we assume that each MS supports  $N_s$  ( $N_s \leq N_{RF}^r$ ) data streams.

Then the total data stream sent by the base station is  $KN_s$  ( $KN_s \leq N_{RF}^t$ ). Subject to the constraints of BS and MS, the number of RF chains are  $KN_s \leq N_{RF}^t \ll N_t$  and  $N_s \leq N_{RF}^r \ll N_r$  respectively. Note that to support  $N_s$  data streams transmission for each user, the least number of RF chains at each MS and the BS are  $N_{RF}^r = N_s$  and  $N_{RF}^t = KN_s = KN_{RF}^r$  respectively. In consideration of hardware cost and power consumption of RF chains, we assume the least number of RF chains in this paper. The number of RF chains at each MS and the BS are  $N_{RF}^r = N_s$  and  $N_{RF}^t = KN_s = KN_{RF}^r$ .

Let  $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_K^T]^T \in \mathbb{C}^{KN_s \times 1}$  be the transmission symbol.  $\mathbf{x}_k \in \mathbb{C}^{N_s \times 1}$  represents the transmitted pilot symbol of the  $k$ th ( $k \in [1, K]$ ) MS, which satisfies  $E[\mathbf{x}_k \mathbf{x}_k^H] = \mathbf{I}_{N_s} P/N_s$ ,  $P$  is the average transmit power,  $\mathbf{I}_{N_s}$  is an identity matrix with dimension  $N_s \times N_s$ . In the BS,  $\mathbf{x}_k$  is first precoded by the digital precoder  $\mathbf{F}_{BB} = [\mathbf{F}_{BB,1}, \mathbf{F}_{BB,2}, \dots, \mathbf{F}_{BB,K}] \in \mathbb{C}^{N_{RF}^t \times KN_s}$  and then precoded by an analog precoder  $\mathbf{F}_{RF} \in \mathbb{C}^{N_{RF}^t \times N_t}$ , which is implemented by an analog phase shifter array. Since the phase shifters in the analog precoder in Fig. 1 can only change the phase of the transmitted signal, each entry of  $\mathbf{F}_{RF}$  is of constant modulus. We normalize its entries to satisfy  $|\mathbf{F}_{RF}| = 1/\sqrt{N_t}$ . In addition, in order to meet the transmit power constraint,  $\mathbf{F}_{BB}$  is normalized to satisfy  $\|\mathbf{F}_{RF} \mathbf{F}_{BB}\|^2 = KN_s$ . After the hybrid precoding, the transmitted signal becomes

$$\mathbf{S}_k = \mathbf{F}_{RF} \mathbf{F}_{BB,k} \mathbf{x}_k. \quad (1)$$

Therefore, the received signal of the  $k$ th user is given by

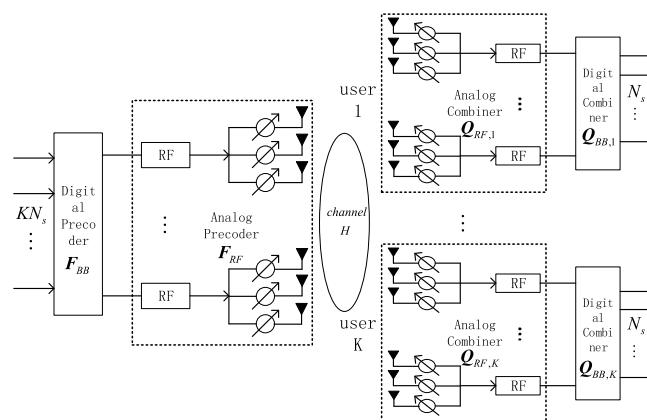
$$\bar{\mathbf{Y}}_k = \mathbf{H}_k \mathbf{S}_k + \mathbf{n}_k, \quad (2)$$

where  $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$  represents multipath fading wideband mmWave channels which are explained in section III-B,  $\mathbf{n}_k \in \mathbb{C}^{N_r \times 1}$  is the additive complex Gaussian noise at the  $k$ th MS and each entry of  $\mathbf{n}_k$  follows the independent and identically distributed (i.i.d.) complex Gaussian distribution with zero mean and variance  $\sigma^2$ .

At the  $k$ th MS, firstly the received signals are processed by a RF combiner  $\mathbf{Q}_{RF,k} \in \mathbb{C}^{N_r \times N_{RF}^r}$  which is implemented by analog phase shifters similar to the BS. We also normalize its entries to satisfy  $|\mathbf{Q}_{RF,k}| = 1/\sqrt{N_r}$ . Then followed by a baseband combiner  $\mathbf{Q}_{BB,k} \in \mathbb{C}^{N_{RF}^r \times N_s}$  to detect its symbol. Then, the detection signal of the  $k$ th MS is given by

$$\begin{aligned} \mathbf{Y}_k &= \mathbf{Q}_{BB,k}^H \mathbf{Q}_{RF,k}^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{F}_{BB,k} \mathbf{x}_k \\ &\quad + \sum_{j=1, j \neq k}^K \mathbf{Q}_{BB,j}^H \mathbf{Q}_{RF,j}^H \mathbf{H}_j \mathbf{F}_{RF} \mathbf{F}_{BB,j} \mathbf{x}_j \\ &\quad + \mathbf{Q}_{BB,k}^H \mathbf{Q}_{RF,k}^H \mathbf{n}_k \end{aligned} \quad (3)$$

where the three consecutive terms respectively represent the  $k$ th user's desired signal, interference signals from other



**FIGURE 1.** The multi-user massive MIMO system.

users, and additive Gaussian noise. By stacking all  $\{\mathbf{Y}_k\}_{k=1}^K$  into a long vector, the multi-user signal becomes

$$\mathbf{Y} = \sum_{n_r=1}^{N_r} \sum_{n_t=1}^{N_t} \sum_{i=1}^K \mathbf{Q}_i^H \mathbf{H}_i \mathbf{F}_i \mathbf{x}_i + \mathbf{Q}_i^H \mathbf{n}_i, \quad (4)$$

where  $\mathbf{Y} \in \mathbb{C}^{N_t \times N_r}$ ,  $\mathbf{Q}_i = \mathbf{Q}_{RF,i}^H \mathbf{Q}_{BB,i}$ ,  $\mathbf{F}_i = \mathbf{F}_{RF} \mathbf{F}_{BB,i}$ .

### B. CHANNEL MODEL

Measurement campaigns in dense-urban Non line of sight (NLOS) environments reveal that mmWave channels typically exhibit limited scattering characteristics [39]. Also, considering the wideband nature of mmWave channels, we adopt a geometric wideband mmWave channel model with  $L$  scatterers between the MS and the BS. The  $l$ th ( $l \in [1, L]$ ) scatterer has a time delay  $\tau_l$ .  $\theta_l \in [0, 2\pi]$  and  $\phi_l \in [0, 2\pi]$  represent azimuth and elevation angles of departure (AoDs) and arrival (AoAs) at the transceiver. Under this model, the  $k$ th MS channel matrix in the delay domain can be written as

$$\mathbf{H}_k = \sum_{l=1}^L \alpha_{k,l} \mathbf{a}_r(\theta_{r,k,l}, \phi_{r,k,l}) \mathbf{a}_t^H(\theta_{t,k,l}, \phi_{t,k,l}) e^{-j2\pi\tau_{k,l}f_s}, \quad (5)$$

where  $\alpha_{k,l}$  is the complex path gain associated with the  $l$ th path.  $\mathbf{a}_r(\theta_{r,k,l}, \phi_{r,k,l}) \in \mathbb{C}^{N_r}$  and  $\mathbf{a}_t(\theta_{t,k,l}, \phi_{t,k,l}) \in \mathbb{C}^{N_t}$  are the antenna array response vectors of the MS and BS, respectively.  $\theta_{r,k,l}$  and  $\phi_{r,k,l}$  are azimuth and elevation AOAs of the  $l$ th path, respectively.  $\theta_{t,k,l}$  and  $\phi_{t,k,l}$  are the azimuth and elevation AODs of the  $l$ th path, respectively.  $f_s$  denotes the sampling rate.

For convenience, (5) can be given by

$$\mathbf{H}_k = \mathbf{E}_k \mathbf{A}_{r,k} \mathbf{A}_{t,k}^H, \quad (6)$$

where

$$\mathbf{E}_k = \text{diag}\left(\alpha_{k,1} e^{-j2\pi\tau_{k,1}f_s}, \dots, \alpha_{k,L} e^{-j2\pi\tau_{k,L}f_s}\right), \quad (7a)$$

$$\mathbf{A}_{r,k} = [\mathbf{a}_r(\theta_{r,k,1}, \phi_{r,k,1}), \dots, \mathbf{a}_r(\theta_{r,k,L}, \phi_{r,k,L})], \quad (7b)$$

$$\mathbf{A}_{t,k} = [\mathbf{a}_t(\theta_{t,k,1}, \phi_{t,k,1}), \dots, \mathbf{a}_t(\theta_{t,k,L}, \phi_{t,k,L})], \quad (7c)$$

The antenna array response vectors of the BS and MS are given by

$$\begin{aligned} \mathbf{a}_t(\theta_t, \phi_t) &= \frac{1}{\sqrt{N_t}} \left[ 1, e^{j(2\pi/\lambda)d \sin \phi_t \cos \theta_t}, \dots, \right. \\ &\quad \left. e^{j(N_t-1)(2\pi/\lambda)d \sin \phi_t \cos \theta_t} \right]^T, \end{aligned} \quad (8a)$$

$$\begin{aligned} \mathbf{a}_r(\theta_r, \phi_r) &= \frac{1}{\sqrt{N_r}} \left[ 1, e^{j(2\pi/\lambda)d \sin \phi_r \cos \theta_r}, \dots, \right. \\ &\quad \left. e^{j(N_r-1)(2\pi/\lambda)d \sin \phi_r \cos \theta_r} \right]^T \end{aligned} \quad (8b)$$

where  $\lambda$  is the carrier wavelength,  $d$  indicates the distance between the antennas.

Our objective is to estimate the channel matrices  $\{\mathbf{H}_i\}_{i=1}^K$  and from the received signals  $\{\mathbf{Y}_i\}_{i=1}^K$  of all users. In particular, we wish to provide a reliable channel estimate by

using as few measurements as possible because the number of measurements is linearly proportional to the number of RF chains at the MS, which is expected to be reduced in practice. The mmWave channel measurement results show that the mmWave has a diffuse scattering phenomenon on the surface of the rough scatterer, and the scattering range will increase as the wavelength decreases [39]. For scenarios where users are dense, when there is not enough space between users, diffuse scattering may cause adjacent users to receive signals of the same path. In practice, this scheme is generally applicable to the scenario where the positions of users are randomly distributed. Users in co-channel deployment may experience significant interference from other users. The channel matrices  $\{\mathbf{H}_i\}_{i=1}^K$  are characterized by a set of common parameters  $\{\alpha_i, \theta_i, \phi_i, \tau_i\}_{i=1}^K$ . It can be known from the above analysis that the channel matrices contain the spatial characteristics of each transmitter, and the azimuth  $\{\theta_i\}_{i=1}^K$ , elevation  $\{\phi_i\}_{i=1}^K$ , fading factors  $\{\alpha_i\}_{i=1}^K$  and correlation delay  $\{\tau_i\}_{i=1}^K$ . Thus it can be expected that a joint estimation of  $\{\mathbf{H}_i\}_{i=1}^K$  improves the estimation accuracy.

Note that the problem of single-user mmWave channel estimation has been studied in [40], [41]. Specifically, to estimate the downlink channel, the BS employs  $M$  different beamforming vectors at  $M$  successive time frames, and at each time frame, the MS uses  $Q$  combining vectors to detect the signal transmitted over each beamforming vector. By exploiting the sparse scattering nature of mmWave channels, the problem of estimating the mmWave channel can be formulated as a sparse signal recovery problem and the training overhead can be considerably reduced. The above method can also be used to solve our uplink channel estimation problem if channels from users to the BS are estimated separately. Nevertheless, we will show that a joint estimation (of multi-user channels) scheme may lead to an additional training overhead reduction.

### IV. SEGMENTATION AND CLUSTERING

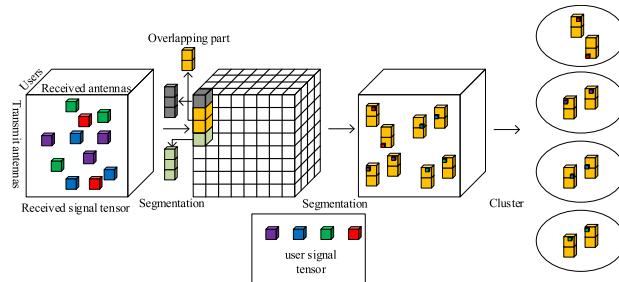
In recent years, tensor has been proved to be an effective analysis tool in multi-user massive MIMO. To solve the problem of high complexity of channel estimation calculation in massive MIMO systems, this paper exploits the tensor strategy to analyze the received signal. We use the two-dimensional spatial and user domain of the signal to analysis and cluster, and then combine similar channel characteristics into sparse and low-rank tensor decomposition models to effectively detect and separate multi-user information of mixed signals. Dictionary learning can express the multi-source localization (MSL) problem as joint parameter dictionary learning (PDL) and sparse signal recovery (SSR). It is also possible to build a dictionary and sparse coding on the sub-band spectrum to improve the signal recognition ability. The wrong channel model can also be embedded in dictionary learning to improve accuracy [42]–[44]. After users are clustered into groups, we build a dictionary of user channel information in the user group. Then we can estimate the user group channel more efficiently.

Substituting the channel representation from (6) into (4), the multi-user signal becomes

$$\mathcal{Y} = \sum_{n_r=1}^{N_r} \sum_{n_t=1}^{N_t} \sum_{i=1}^K \vec{\mathbf{E}}_i \vec{\mathbf{A}}_{r,i} \circ \vec{\mathbf{A}}_{t,i} \circ \mathbf{x}_i + \mathcal{N}_i, \quad (9)$$

where  $\vec{\mathbf{A}}_{r,i} = \mathbf{Q}_i^H \mathbf{A}_{r,i}$ ,  $\vec{\mathbf{A}}_{t,i} = \mathbf{A}_{t,i}^H \mathbf{F}_i$  and  $\mathcal{Y} \in \mathbb{C}^{N_r \times N_t \times K}$  is now in Tucker form [36].  $\mathcal{N}_i \in \mathbb{C}^{N_r \times N_t \times K}$  is additive Gaussian noise tensor.

After constructing the received signals of all users into a tensor form, we perform convolutional segmentation and grouping of the received signals to prepare for the following channel estimation of the user group (Fig. 2).



**FIGURE 2.** Segmentation and clustering of received signals.

#### A. CONVOLUTIONAL SEGMENTATION OF RECEIVED SIGNALS

In the received signal of  $\mathcal{Y}$ , the 3D cube in the form of third-order tensor contains rich multi-dimensional structure information. Because different sources have different spatial characteristics, multi-source multipath grouping can be achieved by estimating the parameters of the spatial characteristics.

The received signal of users can be regarded as three dimensional tensor  $\mathcal{Y} \in \mathbb{C}^{N_r \times N_t \times K}$  of transmitting antenna spatial characteristics, receiving antenna spatial characteristics and users. First, the received signal  $\mathcal{Y}$  is divided into overlapping three-dimensional sub-tensors, where the sliding window step size  $\omega = 1$ . Therefore, in the spatial dimension of the receiving antenna, the receiving antenna array is divided into  $(N_r - \Delta n_r + 1)$  sub-array blocks, and each sub-array element is  $\Delta n_r$ . In the spatial dimension of the transmitting antenna, all transmitting antenna arrays are divided into  $(N_t - \Delta n_t + 1)$  sub-array blocks, each sub-array element is  $\Delta n_t$ . In the user dimension, all users are divided into  $(K - \Delta g + 1)$  groups. By convolutionally dividing the two-dimensional space of the signal tensor, we obtain  $(N_r - \Delta n_r + 1)(N_t - \Delta n_t + 1)$  sub-tensors. Subtensor  $\Delta \mathcal{Y} \in \mathbb{C}^{\Delta n_r \times \Delta n_t \times \Delta g}$  of received signal with multidimensional space information is expressed as:

$$\Delta \mathcal{Y} = \sum_{n_r=p}^{p+\Delta n_r-1} \sum_{n_t=q}^{q+\Delta n_t-1} \sum_{g=1}^{g'+\Delta g-1} \vec{\mathbf{E}}_g \vec{\mathbf{A}}_{n_r,g} \circ \vec{\mathbf{A}}_{n_t,g} \circ \mathbf{x}_g + \mathcal{N}_g, \quad (10)$$

where  $1 \leq p \leq N_r - \Delta n_r$ ,  $1 \leq q \leq N_t - \Delta n_t$ ,  $1 \leq g' \leq K - \Delta g n_r = p, \dots, p + \Delta n_r$ ,  $n_t = p, \dots, p + \Delta n_t$ ,  $g = g', \dots, g' + \Delta g$ .  $\Delta \mathcal{Y}$  represents the subtensor after convolutional segmentation.  $\mathcal{N}_g$  represents white noise tensor.

DL has been effectively applied to multi-user interference canceling and channel denoising by considering the nonlocal similarity property of users [45]. The basic idea is to firstly cluster the similar users into groups. and then to encourage each group share similar atoms in the dictionary. The DL model can be extended to the channel estimation of the user group. First, we construct user group patches for channel estimation of the user group as follows. Receiving signal of all users with  $N_r \times N_t$  spatial dimension and  $K$  users can be expressed as a three-order tensor  $\mathcal{Y} \in \mathbb{C}^{N_r \times N_t \times K}$ . By sweeping all across the receiving signal of all users with overlaps, we can build a user group of patches  $\{\mathcal{Y}^{(p,q,g)}\}_{1 \leq p \leq N_r - \Delta n_r, 1 \leq q \leq N_t - \Delta n_t, 1 \leq g \leq K - \Delta g} \subset \mathcal{Y}^{N_r \times N_t \times K}$  from all users. For simplicity, we reformulate all patches as a group of 3D patches  $\{\mathcal{Y}_i\}_{i=1}^I$ , where  $I = (N_r - \Delta n_r + 1)(N_t - \Delta n_t + 1)(K - \Delta g + 1)$  denotes the patch number. Each patch constructed in this way contains two spatial dimension and a user dimension. This can easily help us consider an important attribute of all users: the relevance of users with similar channels.

Based on this patch set  $\{\mathcal{Y}_i\}_{i=1}^I$ , the user DL model can then be constructed to calculate the spatial and user dictionaries  $\{\vec{\mathbf{A}}_{n_r,g} \in \mathbb{C}^{\Delta n_r \times L}, \vec{\mathbf{A}}_{n_t,g} \in \mathbb{C}^{\Delta n_t \times L}, \mathbf{x}_g \in \mathbb{C}^{\Delta g \times K}\}$ , implying the redundancy of these dictionaries, as follows:

$$\min_{\vec{\mathbf{A}}_{n_r,g}, \vec{\mathbf{A}}_{n_t,g}, \mathbf{x}_g, \vec{\mathcal{E}}_g} \sum_{i=1}^I \left\| \Delta \mathcal{Y} - \vec{\mathcal{E}}_g \times_1 \vec{\mathbf{A}}_{n_r,g} \times_2 \vec{\mathbf{A}}_{n_t,g} \times_3 \mathbf{x}_g \right\|, \quad (11)$$

where  $\vec{\mathcal{E}}_g \in \mathbb{C}^{L \times L \times K}$  corresponds to the coefficient tensor for  $\Delta \mathcal{Y}$  which governs the affiliated interaction between the dictionaries.

#### B. CLUSTERING OF RECEIVED SIGNALS

We use the idea of dictionary learning to estimate the channel of the user group, that is, use the correlation of users with similar channel characteristics. Therefore, users with similar channel characteristics need to be clustered into groups. The established user channel model is based on a geometric stochastic channel model, using the concept of scattering clusters, which contains many stochastically varying multipath components. Scattering clusters can be used to cluster user groups. The starting point is a large number of multidimensional parametric channel estimation data, obtained from MIMO measurements. It has been investigated that these parameters appear in clusters [46]–[48], i.e. in groups of multipath components (MPCs) with similar parameters, such as delay, AoAs and AoDs. We show that using the well-known K-means clustering algorithm [49] with multipath component distance (MCD) as distance measure improves performance considerably.

K-means algorithm identifies each cluster by its centroid position in the parameter space, then it is assumed that there are  $G$  centroids. Each MPC is assigned to the cluster centroid with the smallest distance. The algorithm repeatedly optimizes the positions of the centroids to minimize the total distance from each MPC to its centroid, as:

$$D = \sum_{g=1}^G d(x_g, c_{x_g}), \quad (12)$$

where  $x_g$  denotes the parameter vector of  $g$ th  $g \in [1, \dots, G]$  MPC,  $c_{x_g}$  denotes the parameters of cluster centroid closest to the  $g$ th MPC, and  $d(\cdot)$  represents the distance function between any two points in the parameter space.

**Distance measures:** It was customary [49] to calculate the distance for each dimension (delay, AoAs, AoDs) separately. Clustering was subsequently done either sequentially (first delay domain, then subsequently clustering AoAs and AoDs) or jointly. We demonstrate that joint clustering is more promising as the cluster structure in the data is more visible in high-dimensional spaces, but the data in different dimensions (coming even in different units) has to be scaled. To identify clusters correctly, we use the multipath component distance (MCD) [46]. This metric scales the data to enable joint clustering. To assess the performance, we compare the squared Euclidean distance with the MCD.

The squared Euclidean distance between any two estimated MPCs  $i$  and  $j$  is given by

$$d_{\tau,ij}^2 = (\tau_i - \tau_j)^2. \quad (13)$$

The metric is extended for AoAs and AoDs to cope with the angular periodicity, according to

$$d_{\text{AoA},ij}^2 = p v^2 (\varphi_{\text{AoA},i} - \varphi_{\text{AoA},j}, \pi), \quad (14)$$

where  $p v^2(\cdot, \pi)$  maps  $(\cdot)$  to its principal value in the interval  $[-\pi, \pi]$ . The metric reads similar for the AoDs. This metric is only useful for one-dimensional clustering, as it does not scale the data.

Assuming that subtensor  $(p, q)$ th and subtensor  $(u, v)$ th have  $L^{(p,q)}$  paths and  $L^{(u,v)}$  paths, respectively, the user dimension is kept fixed at this time.

The distance of the AoA for the  $l$ th path is defined as:

$$MCD_{\text{AoA},l}^{(p,q),(u,v)} = \frac{1}{2} \left| \begin{pmatrix} \sin \theta_{r_k,l}^{(p,q)} \cos \phi_{r_k,l}^{(p,q)} \\ \sin \theta_{r_k,l}^{(p,q)} \sin \phi_{r_k,l}^{(p,q)} \\ \cos \phi_{r_k,l}^{(p,q)} \end{pmatrix} - \begin{pmatrix} \sin \theta_{r_k,l}^{(u,v)} \cos \phi_{r_k,l}^{(u,v)} \\ \sin \theta_{r_k,l}^{(u,v)} \sin \phi_{r_k,l}^{(u,v)} \\ \cos \phi_{r_k,l}^{(u,v)} \end{pmatrix} \right| \quad (15)$$

The distance of the AoD for the  $l$ th path is defined as:

$$MCD_{\text{AoD},l}^{(p,q),(u,v)} = \frac{1}{2} \left| \begin{pmatrix} \sin \theta_{t,l}^{(p,q)} \cos \phi_{t,l}^{(p,q)} \\ \sin \theta_{t,l}^{(p,q)} \sin \phi_{t,l}^{(p,q)} \\ \cos \phi_{t,l}^{(p,q)} \end{pmatrix} - \begin{pmatrix} \sin \theta_{t,l}^{(u,v)} \cos \phi_{t,l}^{(u,v)} \\ \sin \theta_{t,l}^{(u,v)} \sin \phi_{t,l}^{(u,v)} \\ \cos \phi_{t,l}^{(u,v)} \end{pmatrix} \right| \quad (16)$$

The delay distance for the  $l$ th path is defined as:

$$MCD_{\tau_l}^{(p,q),(u,v)} = \eta \cdot \frac{|\tau_l^{(p,q)} - \tau_l^{(u,v)}|}{\Delta \tau_l} \cdot \frac{\tau_{sd}}{\Delta \tau_l}, \quad (17)$$

where  $\eta$  is the scale factor that adjusts the delay weight in the distance function.  $\eta$  is a suitable delay scaling factor to give the delay more ‘importance’ when necessary.  $\Delta \tau_l$  is the range of delay, and  $\Delta \tau_l = \max^{(p,q),(u,v)} \{ |\tau_l^{(p,q)} - \tau_l^{(u,v)}| \}$ .  $\tau_{sd}$  is the standard deviation of delay.

According to (15), (16), and (17), the multi-path component distance metric between the two subtensors is (18), as shown at the bottom of the page.

In (18), the data can be scaled appropriately to conform to the MCD metric.  $L$  is the maximum number of paths in the two subchannel tensors, that is  $L = \max(L^{(p,q)}, L^{(u,v)})$ , when  $L^{(p,q)} < L$  or  $L^{(u,v)} < L$ , the missing bits are filled with zero.

Our data set is  $\Phi$  which is the MPC parameter group obtained from the  $(N_r - \Delta n_r + 1)(N_t - \Delta n_t + 1)(K - \Delta g + 1)$  subtensors. The clustering process is as follows:

1. Randomly initialize the group centroid  $\mu_1, \mu_2, \dots, \mu_g \subset \Phi$ , that is independently select  $G$  centroids from the data set  $\Phi$  as the clustering center of the subtensor.

2. Distribute  $\{x_i^{MPC}\}_{i=1}^I$  MPC parameter samples of each subtensor to a reasonable group centroid  $\mu_g$ , for each  $x_i^{MPC}$ :

$$c^{(i)} : \arg \min_g d_{\text{MCD}}(x_i^{MPC}, \mu_g), \quad (19)$$

where  $g = 1, 2, \dots, G$ ,  $c^{(i)}$  represents the set index of the  $i$ th subtensor and  $g$ th group centroid cluster.  $d_{\text{MCD}}(x_i^{MPC}, \mu_g)$  is the multipath component distance between the MPC parameter sample  $x_i^{MPC}$  of  $i$ th subtensor and centroid  $\mu_g$ . For specific solutions, see (15) and (16), (17) and (18).

3. Update the centroid of the cluster. The centroid of the  $g$ th group is update to:

$$\mu_g := \frac{\sum_{i=1}^I \mathbb{1}\{c^{(i)} = g\} \cdot x_i^{MPC}}{\sum_{i=1}^I \mathbb{1}\{c^{(i)} = g\}}, \quad (20)$$

$$MCD^{(p,q),(u,v)} = \sum_l^L \sqrt{\|MCD_{\tau_l}^{(p,q),(u,v)}\|^2 + \|MCD_{\text{AoD},l}^{(p,q),(u,v)}\|^2 + \|MCD_{\text{AoA},l}^{(p,q),(u,v)}\|^2} \quad (18)$$

4. Repeat steps 2 and 3 until convergence, namely the cluster centroid position remains unchanged.

5. Output a set of  $G$  group subtensors.

The K-means algorithm iteratively optimizes the position of group centroid. Its clustering algorithm performance can be measured by the total MPC distance from each subtensor to its centroid, namely the total MCD square sum of all objects in the group to the group centroid, which is defined as:

$$D' = \sum_{i=1}^I d_{\text{MCD}}(x_i^{\text{MPC}}, \mu_{c^{(i)}})^2, \quad (21)$$

where  $d_{\text{MCD}}(x_i^{\text{MPC}}, \mu_{c^{(i)}})$  represents the updated MCD of group centroid  $\mu_{c^{(i)}}$  and objects  $x_i^{\text{MPC}}$ , and  $x_i^{\text{MPC}} \in \mu_{c^{(i)}}$ .

According to the above analysis and processing, users with similar channel characteristics are clustered into user groups. The number of user groups is  $G$ , assuming that the number of users in the  $g$ th  $g \in [1, \dots, G]$  user group is  $M_g$ .

The received signal in  $g$ th group can be given by

$$\tilde{\mathcal{Y}}_g = \tilde{\mathcal{E}}_g \times_1 \tilde{\mathbf{A}}_{n_r,g} \times_2 \tilde{\mathbf{A}}_{n_t,g} \times_3 \tilde{\mathbf{x}}_g + \sum_{g'=1, g' \neq g}^G \tilde{\mathcal{E}}_{g'} \times_1 \tilde{\mathbf{A}}_{n_r,g'} \times_2 \tilde{\mathbf{A}}_{n_t,g'} \times_3 \tilde{\mathbf{x}}_{g'} + \mathcal{N}_g, \quad (22)$$

where three consecutive terms represent the signal tensor of the  $g$ th user group, the signal tensor interference of other user groups, and the additive white Gaussian noise tensor.

After clustering users with similar channel characteristics into groups.  $\{\mathcal{Y}_{i_g}\}_{i=1}^{M_g}$  ( $g = 1, 2, \dots, G$ ) denotes the signals of all users in the group, where  $G$  is the group number,  $M_g$  is the user number in the  $g$ th group and  $i_g$  denotes the  $i$  user in the  $g$ th cluster. And then we attempt to enforce each group share the similar atoms in each of the spatial dictionaries  $\tilde{\mathbf{A}}_{n_r,g}$ ,  $\tilde{\mathbf{A}}_{n_t,g}$  and user pilots  $\tilde{\mathbf{x}}_g$ . For convenience we combine the subtensors in the  $g$ th group together to formulate a four-order tensor:  $\tilde{\mathcal{Y}}_g \in \mathbb{C}^{n_r \times n_t \times G \times M_g}$ . The supplemental fourth dimension corresponds to the number of users in each group.

Analogously, we align all coefficient tensors corresponding to the  $g$ th user group to form  $\tilde{\mathcal{E}}_g \in \mathbb{C}^{n_r \times n_t \times G \times M_g}$ . Then the aim of the user tensor DL can be attained by the following group-block-sparsity regularizer.

**Definition 1(Group-Block-Sparsity):** For a coefficient tensor  $\tilde{\mathcal{E}}_g \in \mathbb{C}^{n_r \times n_t \times G \times M_g}$ , its group-block-sparsity with respect to the spatial and user mode is  $\|\tilde{\mathcal{E}}_g\|_B = (r^{\text{MS}}, r^{\text{BS}}, r^s)$  modes is if and only if the smallest index subsets  $I^{\text{MS}}, I^{\text{BS}}, I^s$  satisfying  $(\tilde{\mathcal{E}}_g)_{i_1 i_2 i_3 i_4} = 0$  for all  $(i_1, i_2, i_3) \notin I^{\text{MS}} \times I^{\text{BS}} \times I^s$  contain  $r^{\text{MS}}, r^{\text{BS}}, r^s$  elements, respectively.  $\text{sub}(\tilde{\mathcal{E}}_g) \in \mathbb{C}^{r^{\text{MS}} \times r^{\text{BS}} \times r^s \times \psi}$  denotes the intrinsic subtensor of  $\tilde{\mathcal{E}}_g$  extracted from the entries of the three dimensions of  $\tilde{\mathcal{E}}_g$  specified by the index sets  $I^{\text{MS}}, I^{\text{BS}}, I^s$ , respectively.

Note that the group-sparsity [45] can be seen as the degenerated case of the group-block-sparsity in user groups. Furthermore, when we set  $\psi = 1$  (meaning only one user in

a cluster), the group-block-sparsity so defined exactly corresponds to the concept of block sparsity proposed in [6], which has been substantiated to be capable of enhancing better recovery of the original high order signals since it implicitly incorporates valuable prior information on real signals and facilitates making full use of the dictionary atoms of each mode in signal representation.

Then we can construct the following DL model:

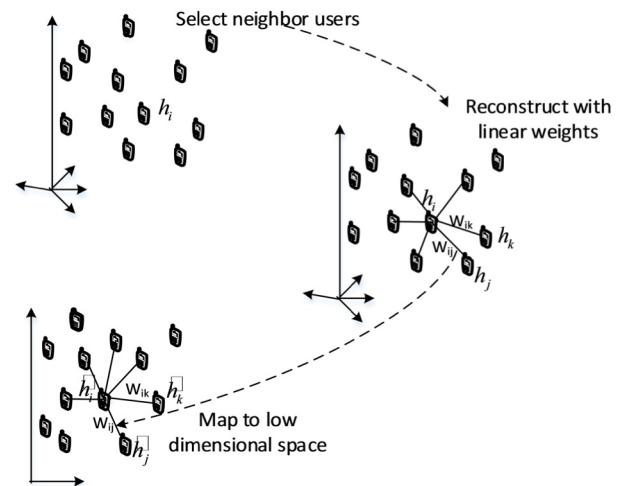
$$\min_{\tilde{\mathcal{A}}_{n_r,g}, \tilde{\mathcal{A}}_{n_t,g}, \tilde{\mathbf{x}}_g} \sum \left\| \tilde{\mathcal{Y}}_g - \tilde{\mathcal{E}}_g \times_1 \tilde{\mathbf{A}}_{n_r,g} \times_2 \tilde{\mathbf{A}}_{n_t,g} \times_3 \tilde{\mathbf{x}}_g \right\|, \quad (23)$$

The group-block-sparsity of  $\tilde{\mathcal{E}}_g$  guarantees that each group  $\tilde{\mathcal{Y}}_g$  shares  $r_g^{\text{MS}}, r_g^{\text{BS}}, r_g^s$  atoms of the dictionaries  $\tilde{\mathbf{A}}_{n_r,g}, \tilde{\mathbf{A}}_{n_t,g}, \tilde{\mathbf{x}}_g$  respectively, and thus the nonlocal similarity among these cluster samples can then be implied.

## V. THE MANIFOLD LEARNING

We know from the previous section that there are  $M_g$  users in  $g$ th group. Since the subtensors with similar channel characteristics in each group are clustered together, the subtensors in the same group have the strongest correlation. But the real relationship between users is not known, and the group tensor dimension at this time is high, which is inconvenient for group tensor channel estimation. Let  $\mathbf{H}_{M_g}$  denote the channel matrix between the BS and the  $M_g$  user in group  $g$ .  $m_g$  ( $1 \leq m_g \leq M_g$ ) denotes the  $m_g$ th user in group  $g$ . Let's denote the matrix corresponding to the  $g$ th cluster  $\tilde{\mathcal{Y}}_g$  as  $\mathbf{H}_g = [\mathbf{H}_1, \dots, \mathbf{H}_{m_g}, \dots, \mathbf{H}_{M_g}] \in \mathbb{C}^{n_r \times n_t \times G \times M_g}$ .

Manifold learning can reduce the dimensionality of high-dimensional data [50]. The basic steps of local linear embedding (LLE) algorithm show that the linear combination relationship can only play a role near the neighborhood [51]. The basic steps of LLE algorithm are illustrated in Fig. 3.



**FIGURE 3.** The basic steps of LLE algorithm.

The high-dimensional receive data can be embedded in subspaces while retaining the geometric properties of the underlying channel manifold. When the low-dimensional

receive data manifold are reconstructed, the subtensors receive data maintain the same local neighbor relationship in the corresponding intrinsic low-dimensional space. So that the high-dimensional receive data can be mapped to the globally unique low-dimensional coordinate system. Finally, the global multi-user high-dimensional receive data nonlinearity is transformed into local linearity to achieve dimensionality reduction. The inter-group interferences existing in (22) can be eliminated by the ADMM scheme [52], which will be explained in Section VI. According to the formula, the signal of any user of the same group can be approximated by the approximate linear combination of other nearby  $M_g - 1$  users. The channel coefficients between the  $M_g$  users of the group,  $\mathbf{H}_{m_g}$  in (24) can be generalized as:

$$\mathbf{H}_{m_g} \approx w_{m_g 1} \mathbf{H}_1 + \cdots + w_{m_g n} \mathbf{H}_n + \cdots + w_{m_g M_g} \mathbf{H}_{M_g}, \quad (24)$$

where  $w_{m_g n}$  is the channel correlation coefficient between the user  $m_g$  and its neighbor  $n$ . In order to calculate the reconstruction weight coefficient  $\{w_{m_g n}\}_{m_g \neq n, n=1}^{M_g}$  between the  $m$ th user and its adjacent user, the linear combination effect is optimal. Each high-dimensional channels  $\mathbf{H}_{m_g}$  and its neighbor high-dimensional channels  $\mathbf{H}_{M_g}$  are located in a linear or nearly linear local neighborhood of the manifold by manifold learning. The high-dimensional channels  $\mathbf{H}_{m_g}$  in the neighborhood can be represented by its neighbor high-dimensional channels  $\mathbf{H}_{M_g}$ . When the low-dimensional channel manifolds are reconstructed, the channels maintain the same local neighbor relationship in the corresponding intrinsic low-dimensional space. The objective function must be constructed and the error value of (24) is minimized:

$$\begin{aligned} \varepsilon(w) &= \min \sum_{m_g=1}^{M_g} \left\| \mathbf{H}_{m_g} - \sum_{n \neq m_g, n \in M_g} w_{m_g n} \mathbf{H}_n \right\|^2 \\ s.t. \quad &\begin{cases} \sum_{n \neq m_g, n \in M_g} w_{m_g n} = 1 \\ \forall n \notin M_g, \quad w_{m_g n} = 0, \end{cases} \end{aligned} \quad (25)$$

where, from the constraint conditions,  $\{0 < w_{m_g n}, w_{m_g n'} < 1 | n, n' \in M_g\}$  are any two weight coefficients corresponding to  $\mathbf{H}_n$  and  $\mathbf{H}_{n'}$  respectively. Let  $\alpha \in \mathbb{R}$ ,  $0 < \alpha < 1$ ,  $\{0 < \alpha w_{m_g n} + (1 - \alpha) w_{m_g n'} < 1 | n, n' \in M_g\}$ . Therefore, the constraint set of the weight coefficient is a convex. If the objective function  $\varepsilon(w)$  satisfies the properties of convex function, the constraint condition is a convex set. In order to achieve dimensionality reduction, the data of the high-dimensional space can be mapped to a lower-dimensional space by using the LLE method.

*Proof:* Assuming any two points on the set  $w_{m_g 1}, w_{m_g 2}, \alpha \in (0, 1)$ , then as shown in (26).

Therefore, the objective function (25) is a convex under the constraint condition. In this case, the solution of the reconstruction weight coefficient can be transformed into a

convex optimization problem, that is, the solution of the least

$$\begin{aligned} &\varepsilon [\alpha w_{m_g 1} + (1 - \alpha) w_{m_g 2}] \\ &= \sum_{m_g=1}^{M_g} \left\| \mathbf{H}_m - [\alpha w_{m_g 1} + (1 - \alpha) w_{m_g 2}] \mathbf{H}_1 \right\|^2 \\ &= \sum_{m_g=1}^{M_g} \left\| \mathbf{H}_{m_g} - \alpha w_{m_g 1} \mathbf{H}_1 - w_{m_g 2} \mathbf{H}_1 + \alpha w_{m_g 2} \mathbf{H}_1 \right\|^2 \\ &= \sum_{m_g=1}^{M_g} \left\| \mathbf{H}_{m_g} - [\alpha (w_{m_g 1} \mathbf{H}_1 + w_{m_g 2} \mathbf{H}_2)] \right\|^2 \\ &\leq \alpha \varepsilon (w_{m_g 1}) + (1 - \alpha) \varepsilon (w_{m_g 2}), \end{aligned} \quad (26)$$

convex optimization problem, that is, the solution of the least square problem under two constraints. Assuming that the channel matrices  $\mathbf{H}_{m_g}, \mathbf{H}_n, \mathbf{H}_{n'}$  of the  $g$ th group are all known, a local covariance matrix  $\tilde{\mathbf{R}}_{n, n'}^m$  is constructed:

$$\tilde{\mathbf{R}}_{n, n'}^m = (\mathbf{H}_{m_g} - \mathbf{H}_n)^T (\mathbf{H}_{m_g} - \mathbf{H}_{n'}), \quad (27)$$

Combined with the constraint  $\sum_{n \in M_g} w_{m_g n} = 1$ , the minimum matrix problem of the objective function is solved. At this time,  $w_{m_g n}$  has a closed solution, which is expressed as:

$$w_{m_g n} = \frac{\sum_{n' \in M_g} (\tilde{\mathbf{R}}_{n, n'}^m)^{-1}}{\sum_{n_1 \in M_g} \sum_{n_2 \in M_g} (\tilde{\mathbf{R}}_{n_1, n_2}^m)^{-1}}, \quad (28)$$

After the LLE dimensionality reduction method, the projection in the low-dimensional space of the input channel matrix  $\mathbf{H}_{m_g}$  of a user  $m_g$  in the group  $g$  is  $\tilde{\mathbf{H}}_{m_g}$ . And  $\mathbf{H}_n, \mathbf{H}_{n'}$  the corresponding projections are  $\tilde{\mathbf{H}}_n, \tilde{\mathbf{H}}_{n'}$ , respectively.  $\tilde{\mathbf{H}}_{m_g}$  can be reconstructed by the linear combination approximation:

$$\tilde{\mathbf{H}}_{m_g} \approx w_{m_g n} \tilde{\mathbf{H}}_n + w_{m_g n'} \tilde{\mathbf{H}}_{n'} + \cdots + w_{m_g M_g} \tilde{\mathbf{H}}_{M_g}, \quad (29)$$

Therefore, the low-dimensional space representation  $\tilde{\mathbf{H}}_{m_g}$  corresponding to  $\mathbf{H}_{m_g}$  can be solved by the following kernel function as (30). In (30),  $\tilde{\mathbf{H}}$  is composed of all  $\tilde{\mathbf{H}}_{m_g}$  column vectors,  $\mathbf{I}$  is an identity matrix with dimension  $M_g \times M_g$ .  $\mathbf{I}_{m_g}$  is the  $m_g$ th column of  $\mathbf{I}$ . And  $\mathbf{w}_{m_g}$  denotes the  $m_g$ th column of the matrix  $\mathbf{W}_g$ , with dimension

$$\begin{aligned} \varepsilon(\tilde{\mathbf{H}}) &= \min \sum_{m_g=1}^{M_g} \left\| \tilde{\mathbf{H}}_{m_g} - \sum_{n \neq m_g, n \in M_g} w_{m_g n} \tilde{\mathbf{H}}_n \right\|^2 \\ &= \min \sum_{m_g=1}^{M_g} \left\| \tilde{\mathbf{H}} (\mathbf{I}_{m_g} - \mathbf{w}_{m_g}) \right\|^2 \\ &= \min \left\{ \text{tr} (\tilde{\mathbf{H}} (\mathbf{I} - \mathbf{W}_g) (\mathbf{I} - \mathbf{W}_g)^T \tilde{\mathbf{H}}^T) \right\} \\ s.t. \quad &\begin{cases} \sum_{m_g=1}^{M_g} \tilde{\mathbf{H}}_{m_g} = 0 \\ \tilde{\mathbf{H}} \tilde{\mathbf{H}}^T = \mathbf{I} \end{cases} \end{aligned} \quad (30)$$

$M_g \times M_g$ , is given by:

$$\mathbf{w}_{m_g} = \begin{bmatrix} w_{m_g 1} \\ w_{m_g 2} \\ \vdots \\ w_{m_g M_g} \end{bmatrix}, \mathbf{W}_g = \begin{bmatrix} w_{11} & w_{21} & \cdots & w_{M_g 1} \\ w_{12} & w_{22} & \cdots & w_{M_g 2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1M_g} & w_{2M_g} & \cdots & w_{M_g M_g} \end{bmatrix}, \quad (31)$$

Define  $\mathbf{B} = (\mathbf{I} - \mathbf{W}_g)(\mathbf{I} - \mathbf{W}_g)^T$ . So that, let  $\lambda_1, \lambda_2, \dots, \lambda_{M_g}$  be the non-zero eigenvalues of the matrix  $\mathbf{B}$  by decomposing the eigenvalue for  $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^T = \lambda\tilde{\mathbf{H}}^T$ . And the eigenvalues are sorted from small to large. The corresponding eigenvectors are  $\tilde{\mathbf{H}}_1^T, \tilde{\mathbf{H}}_2^T, \dots, \tilde{\mathbf{H}}_{M_g}^T$ . To minimize the kernel function error  $\varepsilon(\tilde{\mathbf{H}})$ , the low-dimensional channel matrix of the output  $\tilde{\mathbf{H}}^T$  are consisted of the first  $b$  eigenvector of the matrix  $\mathbf{B}$ .  $\mathbf{B}$  can be obtained through  $(\tilde{\mathbf{H}}^T, \Lambda) = \text{eig}(\mathbf{B})$ , where  $\Lambda$  is the eigenvalue matrix. The main diagonal element of  $\Lambda$  is the eigenvalue of  $\mathbf{B}$ .

Since the relationship among the users in the group is known, then we can reconstruct all the users in the group into a new group tensor, and the  $g$ th group can be written as

$$\hat{\mathcal{Y}}_g = \hat{\mathcal{E}}_g \times_1 \hat{\mathbf{A}}_{n_r, g} \times_2 \hat{\mathbf{A}}_{n_t, g} \times_3 \hat{\mathbf{x}}_g \times_4 \mathbf{W}_g + \sum_{g'=1, g' \neq g}^G \hat{\mathcal{E}}_{g'} \times_1 \hat{\mathbf{A}}_{n_r, g'} \times_2 \hat{\mathbf{A}}_{n_t, g'} \times_3 \hat{\mathbf{x}}_{g'} \times_4 \mathbf{W}_{g'}, \quad (32)$$

where  $\hat{\mathcal{Y}}_g \in \mathbb{C}^{N_t \times N_r \times M_g}$  is the received signal of the  $g$ th group after the correlation coefficient of the channel matrix is known.

## VI. CHANNEL ESTIMATION OF THE USER GROUPS

The tensor analysis method is used to analyze all users at the same time to improve work efficiency. But from (32), it can be seen that the desired signal of user groups is still interfered by other group tensors, so we use the tensor ADMM method to eliminate it. Then the tensor-MUSIC method is used to get our estimated channel parameters of user groups.

### A. THE ADMM FOR ELIMINATING INTERFERENCE

For the convenience of calculation, we temporarily simplify (32) as follows

$$\hat{\mathcal{Y}} = \mathcal{X} + \mathcal{Z} + \mathcal{N}, \quad (33)$$

where

$$\hat{\mathcal{Y}} = \hat{\mathcal{Y}}_g \in \mathbb{C}^{N_t \times N_r \times M_g}, \quad (34a)$$

$$\mathcal{X} = \hat{\mathcal{E}}_g \times_1 \hat{\mathbf{A}}_{n_r, g} \times_2 \hat{\mathbf{A}}_{n_t, g} \times_3 \hat{\mathbf{x}}_g \times_4 \mathbf{W}_g, \quad (34b)$$

$$\mathcal{Z} = \sum_{g'=1, g' \neq g}^G \hat{\mathcal{E}}_{g'} \times_1 \hat{\mathbf{A}}_{n_r, g'} \times_2 \hat{\mathbf{A}}_{n_t, g'} \times_3 \hat{\mathbf{x}}_{g'} \times_4 \mathbf{W}_{g'}, \quad (34c)$$

$$\mathcal{N} = \mathcal{N}_g. \quad (34d)$$

$\mathcal{X}, \mathcal{Z}, \mathcal{N}$  represent the desired signal of  $g$ th user group, interference signals from other user groups and additive Gaussian white noise.

Our object is to obtain the desired signal of  $g$ th user group from (33). Without considering Gaussian noise, Lu et.al. [53] used tensor robust principal component analysis to restore low-rank tensors through convex optimization, is given by

$$\begin{aligned} & \min_{\mathcal{X}, \mathcal{Z}} \|\mathcal{X}\|_* + \rho_2 \|\mathcal{Z}\|_1 \\ & \text{s.t. } \mathcal{Y}' = \mathcal{X} + \mathcal{Z}, \end{aligned} \quad (35)$$

where  $\rho_2 = 1/\sqrt{\max(N_t, N_r) M_g}$  [54], (35) can be solved by polynomial-time algorithms, e.g., the standard ADMM.

We follow this convex optimization scheme to eliminate interference from other user groups and noise. In this paper, the eliminating interference is formulated as

$$\begin{aligned} & \min_{\mathcal{X}, \mathcal{Z}, \mathcal{N}} \|\mathcal{X}\|_* + \lambda_1 \|\mathcal{N}\|_F^2 + \lambda_2 \|\mathcal{Z}\|_1 \\ & \text{s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{N} + \mathcal{Z}, \end{aligned} \quad (36)$$

where  $\lambda_1, \lambda_2$  are the optimized values to achieve a satisfactory recovery result of the user group signal.

The constrained problem defined in (36) can be addressed by a quadratic penalty approach, i.e., by solving

$$\begin{aligned} & \min_{\mathcal{X}, \mathcal{Z}, \mathcal{N}} \|\mathcal{X}\|_* + \lambda_1 \|\mathcal{N}\|_F^2 + \lambda_2 \|\mathcal{Z}\|_1 \\ & + \frac{\beta}{2} \|\mathcal{Y} - (\mathcal{X} + \mathcal{N} + \mathcal{Z})\|_F^2, \end{aligned} \quad (37)$$

where  $\beta$  is a factor that guarantees the convergence of the algorithm.

The solution of (37) can be approached by alternating this minimization with respect variables  $\mathcal{X}, \mathcal{N}$  and  $\mathcal{Z}$ . However, the intermediate minimization becomes increasingly ill-posed when  $\beta$  becomes large. The augmented Lagrangian method (ALM) provides another term to mimic Lagrange multiplier and to overcome the ill-posed problem caused by large value of  $\beta$ . The augmented Lagrangian function for problem defined in (36) is

$$\begin{aligned} & \mathcal{L}(\mathcal{X}, \mathcal{Z}, \mathcal{N}, \Lambda_1; \lambda_1, \lambda_2, \beta) \\ & = \|\mathcal{X}\|_* + \lambda_1 \|\mathcal{N}\|_F^2 + \lambda_2 \|\mathcal{Z}\|_1 \\ & + \langle \Lambda_1, \mathcal{Y} - (\mathcal{X} + \mathcal{N} + \mathcal{Z}) \rangle \\ & + \frac{\beta}{2} \|\mathcal{Y} - (\mathcal{X} + \mathcal{N} + \mathcal{Z})\|_F^2, \end{aligned} \quad (38)$$

where  $\Lambda_1$  is the Lagrangian multipliers. ALM is used to minimize  $\mathcal{L}(\cdot)$  with respect to variables  $\mathcal{X}, \mathcal{N}$ , and  $\mathcal{Z}$  while keeping  $\Lambda_1$  fixed. It then updates  $\Lambda_1$  according to the following rule:

$$\Lambda_1^{\sigma+1} \leftarrow \Lambda_1^\sigma + \beta \left( \mathcal{Y} - \left( \mathcal{X}^{\sigma+1} + \mathcal{N}^{\sigma+1} + \mathcal{Z}^{\sigma+1} \right) \right). \quad (39)$$

The ADMM method uses partial updates by keeping other variables fixed each time. Using the ADMM framework for (38), we can update the variables  $\mathcal{X}, \mathcal{N}$ , and  $\mathcal{Z}$  in the

$(\sigma + 1)$ th iteration, by alternately minimizing the following function while keeping the value of  $\Lambda_1$  fixed at  $\Lambda_1^\sigma$

$$\begin{aligned} \mathcal{L}(\mathcal{X}, \mathcal{Z}, \mathcal{N}, \Lambda_1^\sigma; \lambda_1, \lambda_2, \beta) \\ = \|\mathcal{X}\|_* + \lambda_1 \|\mathcal{N}\|_F^2 + \lambda_2 \|\mathcal{Z}\|_1 \\ + \frac{\beta}{2} \left\| \mathcal{Y} - (\mathcal{X} + \mathcal{N} + \mathcal{Z}) + \frac{\Lambda_1^\sigma}{\beta} \right\|_F^2. \end{aligned} \quad (40)$$

In the  $(\sigma + 1)$ th iteration,  $\mathcal{X}^{\sigma+1}$ ,  $\mathcal{N}^{\sigma+1}$ ,  $\mathcal{Z}^{\sigma+1}$  and  $\Lambda_1^{\sigma+1}$  are updated as follows:

For the updating of  $\mathcal{X}^{\sigma+1}$ , we have

$$\begin{aligned} \mathcal{X}^{\sigma+1} &:= \arg \min_{\mathcal{X}} \bar{\mathcal{L}}(\mathcal{X}, \mathcal{N}^\sigma, \mathcal{Z}^\sigma, \Lambda_1^\sigma) \\ &= \arg \min_{\mathcal{X}} \left( \|\mathcal{X}\|_* + \frac{\beta}{2} \left\| \mathcal{Y} - (\mathcal{X} + \mathcal{N}^\sigma + \mathcal{Z}^\sigma) + \frac{\Lambda_1^\sigma}{\beta} \right\|_F^2 \right). \end{aligned} \quad (41)$$

The updating of  $\mathcal{X}^{\sigma+1}$  has a closed-form solution.

For the updating of  $\mathcal{Z}^{\sigma+1}$ , we have

$$\begin{aligned} \mathcal{Z}^{\sigma+1} &:= \arg \min_{\mathcal{Z}} \bar{\mathcal{L}}(\mathcal{X}^{\sigma+1}, \mathcal{N}^\sigma, \mathcal{Z}, \Lambda_1^\sigma) \\ &= \arg \min_{\mathcal{Z}} \left( \lambda_2 \|\mathcal{Z}\|_1 \right. \\ &\quad \left. + \frac{\beta}{2} \left\| \mathcal{Y} - (\mathcal{X}^{\sigma+1} + \mathcal{N}^\sigma + \mathcal{Z}) + \frac{\Lambda_1^\sigma}{\beta} \right\|_F^2 \right) \end{aligned} \quad (42)$$

Equation (42) can be solved by a soft-shinkage operator

$$\mathcal{Z}^{\sigma+1} := \text{shrink} \left( \mathcal{Y} - (\mathcal{X}^{\sigma+1} + \mathcal{N}^\sigma) + \frac{\Lambda_1^\sigma}{\beta}, \frac{\lambda_2}{\beta} \right) \quad (43)$$

where  $\text{shrink}(\cdot, \cdot)$  is an elementwise soft-shrinkage operator. For each element  $a$  of tensor  $\mathcal{Y} - (\mathcal{X}^{\sigma+1} + \mathcal{N}^\sigma) + \frac{\Lambda_1^\sigma}{\beta}$ , we have

$$\text{shrink} \left( a, \frac{\lambda_2}{\beta} \right) \begin{cases} a - \frac{\lambda_2}{\beta}, & a > \frac{\lambda_2}{\beta} \\ 0, & |a| \leq \frac{\lambda_2}{\beta} \\ a - \frac{\lambda_2}{\beta}, & a < -\frac{\lambda_2}{\beta}. \end{cases} \quad (44)$$

The soft-shrinkage operator  $\text{shrink}(\cdot, \cdot)$  is the proximity operator of the  $l_1$ -norm. For  $\mathcal{N}^{\sigma+1}$ , we have

$$\begin{aligned} \mathcal{N}^{\sigma+1} &:= \arg \min_{\mathcal{N}} \bar{\mathcal{L}}(\mathcal{X}^{\sigma+1}, \mathcal{N}, \mathcal{Z}^{\sigma+1}, \Lambda_1^\sigma) \\ &= \arg \min_{\mathcal{N}} \left( \lambda_1 \|\mathcal{N}\|_F^2 \right. \\ &\quad \left. + \frac{\beta}{2} \left\| \mathcal{Y} - (\mathcal{X}^{\sigma+1} + \mathcal{N} + \mathcal{Z}^{\sigma+1}) + \frac{\Lambda_1^\sigma}{\beta} \right\|_F^2 \right). \end{aligned} \quad (45)$$

Similarly, we can obtain the closed-form solution for the updating of  $\mathcal{N}^{\sigma+1}$

$$\mathcal{N}^{\sigma+1} := \left( \frac{\beta}{2\lambda_1 + \beta} \right) \left( \mathcal{Y} + \frac{\Lambda_1^\sigma}{\beta} - (\mathcal{X}^{\sigma+1} + \mathcal{Z}^{\sigma+1}) \right). \quad (46)$$

The update of  $\Lambda_1^{\sigma+1}$  can be formulated as

$$\Lambda_1^{\sigma+1} \leftarrow \Lambda_1^\sigma + \beta \left( \mathcal{Y} - \left( \mathcal{X}^{\sigma+1} + \mathcal{N}^{\sigma+1} + \mathcal{Z}^{\sigma+1} \right) \right). \quad (47)$$

With a fixed  $\beta$ , ADMM converges slowly. Consequently, an adaptive updating strategy for the penalty parameter is adopted

$$\beta_{\sigma+1} = \min(\beta_{\max}, \eta \beta_\sigma), \quad (48)$$

where  $\beta_{\max}$  is the upper bound for  $\beta$ , and  $\eta$  is the adaptive parameter.

## B. THE STOPPING CRITERION FOR ADMM IS

$$\min \left\{ \beta \left( \left\| \mathcal{Z}^\sigma - \mathcal{Z}^{\sigma+1} \right\| + \left\| \mathcal{N}^\sigma - \mathcal{N}^{\sigma+1} \right\| \right), \frac{1}{\beta} \left\| \Lambda_1^{\sigma+1} - \Lambda_1^\sigma \right\| \right\} \leq \varsigma, \quad (49)$$

where,  $\varsigma$  is an infinitesimal number, e.g.,  $10^{-6}$ .

After the ADMM method, interference signals from other user group tensors have been eliminated. The desired signal of  $g$ th user group is given by

$$\bar{\mathcal{Y}}_g = \hat{\mathcal{E}}_g \times_1 \hat{\mathbf{A}}_{n_r, g} \times_2 \hat{\mathbf{A}}_{n_t, g} \times_3 \hat{\mathbf{x}}_g \times_4 \mathbf{W}_g + \mathcal{N}_g. \quad (50)$$

## C. THE CHANNEL MATRIX AND SIGNAL MATRIX

After the tensor ADMM method, interference signals from other user group tensors have been eliminated.

Due to the sparsity of the antenna sub-array block, the ULA antenna has higher resolvability. In this paper, the spatial-multiple signal classification (SMUSIC) of spatial-multiple signal classification (SMUSIC) combined with angle delay estimation is derived by deriving the time and space covariance matrix. And Time-Multiple Signal Classification (TMUSIC) algorithm [55], using the orthogonality between the signal and noise subspace, MUSIC needs to perform a full-spectrum search in space and time to accurately obtain the path Angle and delay information. The actual wireless communication system has a pilot sequence embedded in its frame, which is used as a reference symbol for channel estimation, the pilot symbol occupies some subcarriers, and the rest is used for data transmission. In order to obtain the best estimate, it is assumed that each user in the network sends an orthogonal training symbol set.

According to the channel model in section III.  $\bar{\mathcal{Y}}_g$  in (50) is given by

$$\begin{aligned} \bar{\mathcal{Y}}_g &= \sum_{l=1}^L \hat{\alpha}_{g,l} \hat{\mathbf{Q}}_{r,g,l}(\theta_{r,g,l}, \phi_{r,g,l}) \circ \hat{\mathbf{F}}_{t,g,l}(\theta_{t,g,l}, \phi_{t,g,l}) \\ &\quad \circ e^{-j2\pi\tau_{g,l}f_s} \hat{\mathbf{x}}_g \mathbf{W}_g + \mathcal{N}_g \end{aligned} \quad (51)$$

where

$$\hat{\mathbf{Q}}_{r,g,l}(\theta_{r,g,l}, \phi_{r,g,l}) = \mathbf{Q}_g^H \hat{\mathbf{a}}_r(\theta_{r,g,l}, \phi_{r,g,l}), \quad (52a)$$

$$\hat{\mathbf{F}}_{t,g,l}(\theta_{t,g,l}, \phi_{t,g,l}) = \hat{\mathbf{a}}_t^H(\theta_{t,g,l}, \phi_{t,g,l}) \mathbf{F}_g. \quad (52b)$$

When the MUSIC algorithm is used to jointly estimate the AoA and AoD of MPC, the steering vector involving multiple parameters can be expressed as

$$\begin{aligned} \hat{\mathbf{C}}_g(\theta_{r,g,l}, \phi_{r,g,l}, \theta_{t,g,l}, \phi_{t,g,l}) \\ = \hat{\mathbf{Q}}_{r,g,l}(\theta_{r,g,l}, \phi_{r,g,l}) \circ \hat{\mathbf{F}}_{t,g,l}(\theta_{t,g,l}, \phi_{t,g,l}). \end{aligned} \quad (53)$$

The delay component in (51) is given by

$$\tilde{\mathbf{G}}(\tau_{g,l}) = \hat{g}(\tau_{g,l}) \operatorname{diag}(\hat{\mathbf{x}}_g \mathbf{W}_g), \quad (54)$$

where

$$\hat{g}(\tau_{g,l}) = e^{-j2\pi\tau_{g,l}f_s}. \quad (55)$$

At this point, (51) can be rewritten as

$$\bar{\mathcal{Y}}_g = \vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g}) \vec{\mathbf{D}}_g(\tilde{\mathbf{G}}_g(\tau))^T, \quad (56)$$

where  $\tilde{\mathbf{G}}_g(\tau) = [\tilde{\mathbf{G}}(\tau_{g,1}), \dots, \tilde{\mathbf{G}}(\tau_{g,l}), \dots, \tilde{\mathbf{G}}(\tau_{g,L})]$ .  $\tilde{\mathbf{G}}(\tau_{g,l})$  is the time array vector.

$$\vec{\mathbf{D}}_g = \operatorname{diag}(\vec{\alpha}_{g,1}, \dots, \vec{\alpha}_{g,l}, \dots, \vec{\alpha}_{g,L}), \quad (57)$$

$$\begin{aligned} \vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g}) \\ = [\hat{\mathbf{C}}_g(\theta_{r,g,1}, \phi_{r,g,1}, \theta_{t,g,1}, \phi_{t,g,1}), \dots, \\ \hat{\mathbf{C}}_g(\theta_{r,g,L}, \phi_{r,g,L}, \theta_{t,g,L}, \phi_{t,g,L})]. \end{aligned} \quad (58)$$

Fading rays are generally considered to be uncorrelated in wireless communication applications, and their fading amplitude is assumed to be a zero-mean complex Gaussian distribution. Therefore, the covariance matrix of the fading vector  $\vec{\alpha} = [\vec{\alpha}_{g,1}, \dots, \vec{\alpha}_{g,l}, \dots, \vec{\alpha}_{g,L}]^T$  is:

$$E[\vec{\alpha}\vec{\alpha}^H] = \operatorname{diag}(\sigma_1^2, \dots, \sigma_l^2, \dots, \sigma_L^2), \quad (59)$$

where  $E[\vec{\alpha}\vec{\alpha}^T] = 0$ ,  $\sigma_l^2$  is the average signal power of the  $l$ th path.

In practical applications, according to (59), the spatial covariance matrix corresponding to the tensor can be expressed as follows:

$$\mathbf{R}_s = E[\bar{\mathcal{Y}}_g(\bar{\mathcal{Y}}_g)^H]$$

$$\begin{aligned} &= \vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g}) \\ &\times P_g(\vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g}))^H, \end{aligned} \quad (60)$$

where  $P_g = E[\vec{\mathbf{D}}_g(\tilde{\mathbf{G}}_g(\tau))(\vec{\mathbf{D}}_g(\tilde{\mathbf{G}}_g(\tau)))^H]$ . In (60), the spatial covariance matrix  $\mathbf{R}_s$  and spatial feature matrix  $\vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g})$  share the same column space.

Similarly, the covariance matrix  $\mathbf{R}_t$  of the delay domain of tensor is:

$$\mathbf{R}_t = E[(\bar{\mathcal{Y}}_g)^T (\bar{\mathcal{Y}}_g)^*] = \tilde{\mathbf{G}}_g(\tau) P_g(\tilde{\mathbf{G}}_g(\tau))^H, \quad (61)$$

In (61), after excluding the noise subspace, the delay domain covariance matrix  $\mathbf{R}_t$  and the delay domain feature matrix  $\tilde{\mathbf{G}}(\tau)$  share the same column space.

Eigenvalue decomposition of the spatial covariance matrix  $\mathbf{R}_s$  and the time-domain covariance matrix  $\mathbf{R}_t$  respectively:

$$\mathbf{R}_s = \mathbf{U}_{s,x} \Lambda_{s,x} (\mathbf{U}_{s,x})^H + + \mathbf{U}_{s,n} \Lambda_{s,n} (\mathbf{U}_{s,n})^H, \quad (62)$$

$$\mathbf{R}_t = \mathbf{U}_{t,x} \Lambda_{t,x} (\mathbf{U}_{t,x})^H + + \mathbf{U}_{t,n} \Lambda_{t,n} (\mathbf{U}_{t,n})^H, \quad (63)$$

where the column vectors  $\mathbf{R}_s$  of  $\mathbf{U}_{s,x}$  are the eigenvectors of the signal subspace in the space domain, the column vectors  $\mathbf{R}_t$  of  $\mathbf{U}_{t,x}$  are the eigenvectors of the signal subspace in the space domain.  $\mathbf{U}_{s,n}$  and  $\mathbf{U}_{t,n}$  are the noise subspaces supplemented orthogonally by  $\mathbf{U}_{s,x}$  and  $\mathbf{U}_{t,x}$  column spaces, respectively.

Using the orthogonality of the signal subspace and noise subspace in the space and delay domain, the zero-spectral functions of S-MUSIC and T-MUSIC are respectively defined as (64) and (65), shown at the bottom of the page.

Estimate the angle and delay information by searching the peaks of the following spatial and delay spectra (66) and (67), as shown at the bottom of the page, where  $\mathbf{P}_{s,n} = \mathbf{U}_{s,n}(\mathbf{U}_{s,n})^H$ ,  $\mathbf{P}_{t,n} = \mathbf{U}_{t,n}(\mathbf{U}_{t,n})^H$ .

## VII. NUMERICAL RESULTS

### A. COMPUTATIONAL COMPLEXITY

In our algorithm, the complexity is dominated by the construction of  $\vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g})$  and  $\tilde{\mathbf{G}}_g(\tau)$ . Our algorithm is a gridless search with the complexity of

$$\mathbf{P}_s(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g}) = \frac{1}{(\vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g}))^H \mathbf{U}_{s,n}(\mathbf{U}_{s,n})^H \vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g})}, \quad (64)$$

$$\mathbf{P}_t(\tau) = \frac{1}{(\tilde{\mathbf{G}}_g(\tau))^H \mathbf{U}_{t,n}(\mathbf{U}_{t,n})^H \tilde{\mathbf{G}}_g(\tau)}. \quad (65)$$

$$(\hat{\theta}_{r,g}, \hat{\phi}_{r,g}, \hat{\theta}_{t,g}, \hat{\phi}_{t,g}) = \arg \max_{(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g})} \frac{1}{(\vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g}))^H \mathbf{P}_{s,n} \vec{\mathbf{C}}_g(\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g})}, \quad (66)$$

$$(\hat{\tau}_g) = \arg \max_{(\tau)} \frac{1}{(\tilde{\mathbf{G}}_g(\tau))^H \mathbf{P}_{t,n} \tilde{\mathbf{G}}_g(\tau)} \quad (67)$$

$\mathcal{O}(N_t N_r K G)$ . As a comparison, the computational complexity of the IP-OMP [17] is of order  $\mathcal{O}(N_t N_r K + N_x N_y N_z)$  where  $N_x N_y N_z$  denotes the number of columns of the overcomplete dictionary. The basic algorithm is the Alternating Least Squares (ALS) in MSVD [32], which has a complexity of the order  $\mathcal{O}(3L^2G + L^3(N_t + N_r + K))$  per iteration. The computational complexity of LRTD [31] is  $\mathcal{O}(N_t N_r K (N_x + N_y + N_z))$ . In each iteration, the computational complexity of BCD-R [34] is  $\mathcal{O}\left(q\left(\sum_{n=1}^N \prod_{k=1}^3 3I_k(L^n)^2\right) + \sum_{n=1}^N \sum_{k=1}^3 (L^n)^3\right)$  where  $q$  is the number of iterations at convergence.

## B. SIMULATION RESULTS

In this section, we present simulation results to examine the estimation performance of the proposed scheme. We use the antenna distance of the conventional antenna array. The distance between neighboring antenna elements is assumed to be half the wavelength of the signal. Also, in our experiments, the precoding matrix  $F$  and the combining matrix  $Q$  are randomly generated with their entries uniformly chosen from a unit circle. Typical values are set as follows: the antenna array sizes for symbol transmission are set as  $N_t = 64$ ,  $N_r = 32$ , the number of RF chains are set as  $N_{RF}^t = 12$ ,  $N_{RF}^r = 6$ , the number of paths is set equal to  $L = 4$ ; the number of subcarriers for transmitting pilot symbols is set as  $K_0 = 128$ ; the sampling rate is set to  $f_s = 0.32$ GHz. All simulation results represent an averaged over 5000 independent Monte Carlo runs. The MATLAB Tensor Toolbox [56] is used.

We first examine the estimation accuracy of the channel parameters  $\{\theta_{r,g}, \phi_{r,g}, \theta_{t,g}, \phi_{t,g}, \tau_g\}$  by the mean square error (MSE). To provide a benchmark, we derive the CRB of each parameter as a lower bound of unbiased estimators [57]. Fig. 4, Fig. 5 and Fig. 6 plot the MSE and CRB curves versus the receiving signal-to-noise ratio (SNR), and SNR is given by

$$\text{SNR} = \frac{\|\mathcal{Y}_g - \mathcal{N}_g\|_F^2}{\|\mathcal{N}_g\|_F^2}, \quad (64)$$

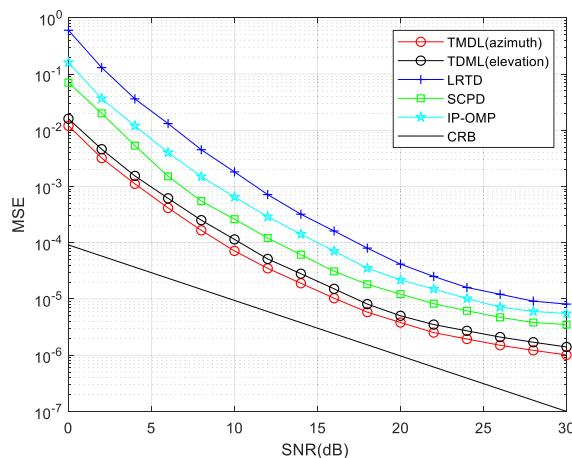


FIGURE 4. MSE performance of AoA vs. SNR.

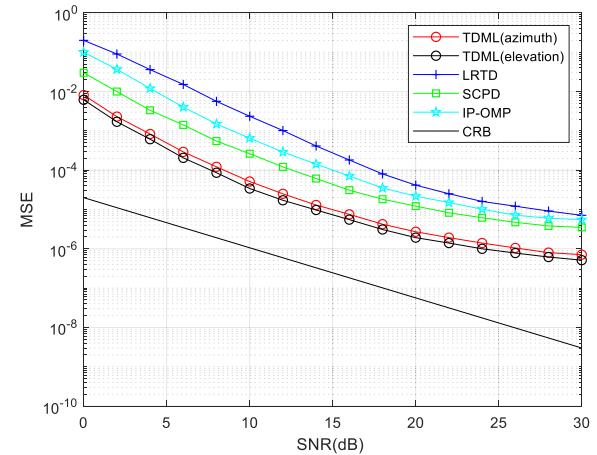


FIGURE 5. MSE performance of AoD vs. SNR.

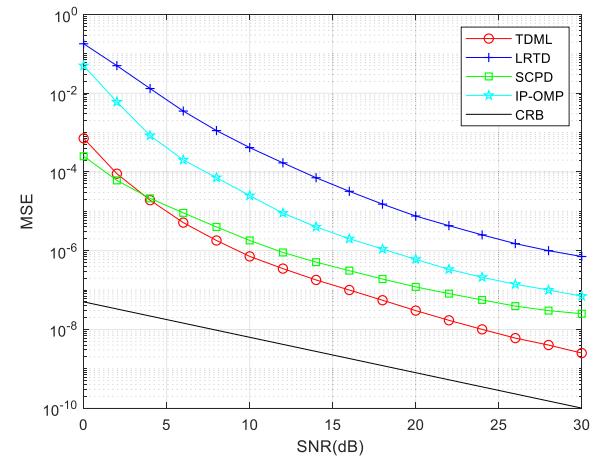


FIGURE 6. MSE performance of delay vs. SNR.

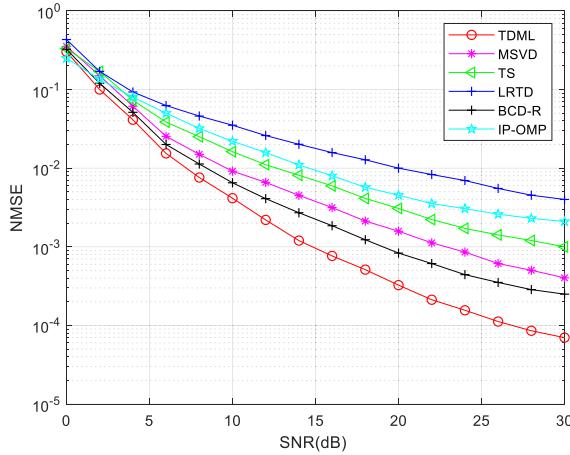
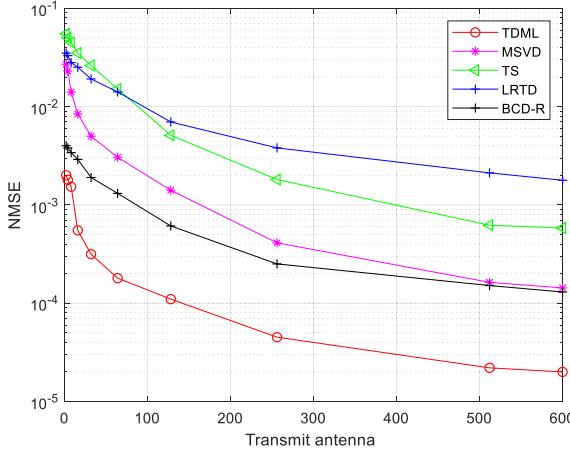
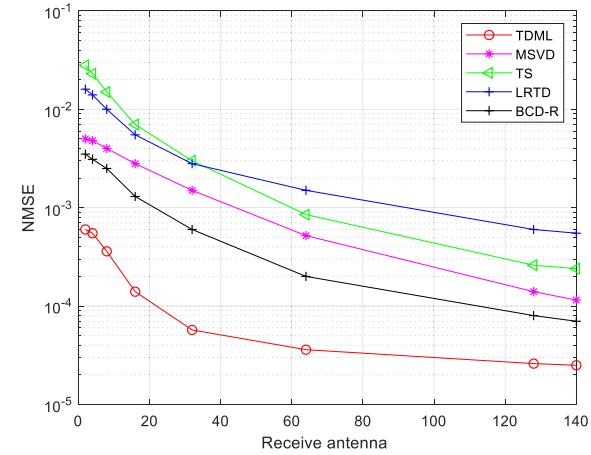
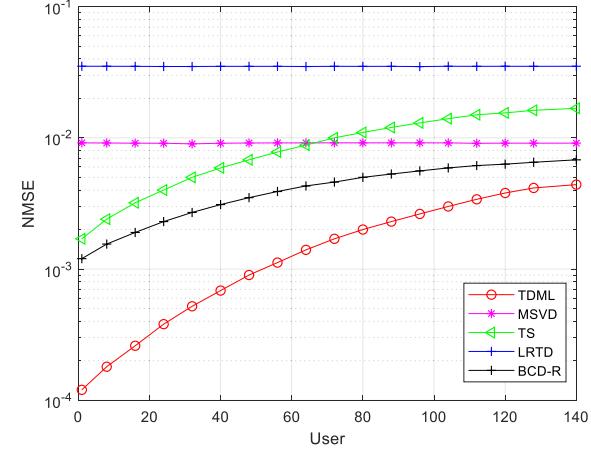
The proposed method is compared with IP-OMP [17], LRTD [31] and SCPD [33]. Our method is abbreviated as TDML. It indicates that the performance of TDML continues to improve as SNR increases. Moreover, TDML consistently outperforms other algorithms, especially in terms of AoA and AoD. In terms of the estimation accuracy of time delay, the accuracy is slightly lower than SCPD at low signal-to-noise ratio. But as the SNR increases, the accuracy is better than SCPD.

Then, we focus on the overall estimation accuracy of channel matrices measured by the normalized MSE (NMSE) and NMSE is defined as

$$NMSE = \frac{1}{R} \left( \sum_{i=1}^R \frac{\|\mathbf{H}_g - \hat{\mathbf{H}}_g\|_F^2}{\|\mathbf{H}_g\|_F^2} \right), \quad (65)$$

where  $\mathbf{H}_g$  denotes the channel matrix associated with the user group, and  $\hat{\mathbf{H}}_g$  is its estimate.  $R$  is number of monte carlo trials.

In Fig. 7, our proposed method is compared with IP-OMP [17], MSVD [32], BCD-R [34], LRTD [31], and TS [35].

**FIGURE 7.** Channel NMSE performance vs. SNR.**FIGURE 8.** Channel NMSE performance vs. Transmit antenna.**FIGURE 9.** Channel NMSE performance vs. Receive antenna.**FIGURE 10.** Channel NMSE performance vs. User.

we set  $N_t = 64$ ,  $N_r = 32$ , and the user  $K = 128$ . The results show that as the SNR increases, the performance of TDML has been improved and is better than other algorithms. This is mainly because other algorithms are limited by the complexity of multi-user high-dimensional channels. BCD-R and IP-OMP use a non-standard tensor decomposition model and the grid search method, so their performance is slightly lower than our solution. Our algorithm clusters users with similar channel characteristics into user groups, and uses manifold learning to reduce dimensionality. Therefore, the performance is higher than other algorithms.

Fig. 8 plots the NMSE curves versus the number of transmit antennas. Fig. 9 plots the NMSE curves versus the number of receive antennas. The SNR is set to 30dB, and the user  $K$  is set to 128. It illustrates all the other methods can achieve enhanced estimation accuracy against the increasing antennas. With the increase of antennas, the beamforming of other algorithms and the reception of MS become better. Their performance is gradually improving. However, the dictionary constructed by our proposed algorithm after convolutional segmentation and clustering is more complete.

This is mainly due to the fact that the two dimensions of the tensor model we build are the transmitting antenna and the receiving antenna. With the increase of antennas, our algorithm uses a large number of antennas for multiple low-dimensional tensor dictionary manifolds after the segmentation of the antennas and the subsequent clustering of similar channels and the learning of manifolds. This allows our algorithm to improve its performance as the number of antennas increases. The accuracy of channel estimation is better than other algorithms.

In Fig. 10, we plot the NMSEs of respective methods vs. and the user  $K$ , where the SNR is set to 10dB. With the increase of users, the performance of LRTD and MSVD hardly changes. The performance of TS and BCD-R continues to decline. With the increase of users, the performance of our solution gradually declines. But it is always better than other solutions. When the number of users reaches a certain level, our algorithm stabilizes. This can be explained by the fact that our algorithm is user group specific and uses a manifold learning approach for dimensionality reduction of high dimensional channels. The performance of other algorithms

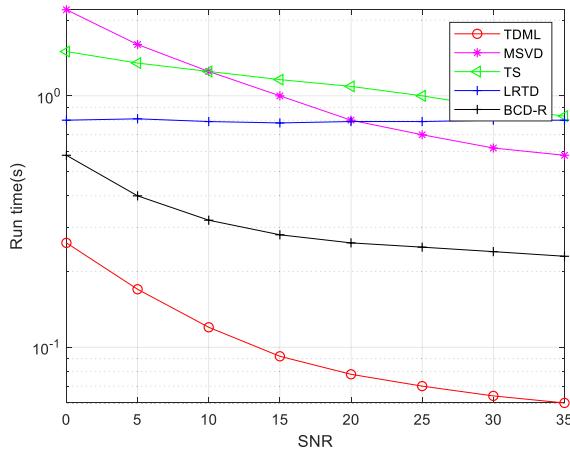


FIGURE 11. Run time vs. SNR.

is limited mainly by the nonlinear low resolution of the high-dimensional channel for multiple users as the number of users increases. In a dense area of users, there are more similar channels between users, and our performance is better.

In Fig. 11 the comparison of the computational complexity is performed by comparing the run time of different algorithms. The user  $K$  is set to 128. The run time is chosen because iterative algorithms are used to achieve a tensor decomposition in MSVD, TS, LRTD, and BCD-R. As seen from the figure, the proposed TDML channel estimation algorithm has the lowest complexity. TS and MSVD yield a much higher complexity especially in the low to medium SNR regime because the tensor decomposition converges slowly.

## VIII. CONCLUSION

In this paper, we propose a tensor dictionary manifold learning method for channel estimation and interference elimination of the multi-user mmWave massive MIMO system. This method uses user groups with similar channels to build tensor manifold dictionary learning of channel information. Specifically, the multi-user signal at the user end is established as a third-order tensor. We prove that users with similar channel characteristics can build a tensor pop dictionary of user group channels. Research on the extraction of channel information shows that even a few pilots can obtain the time delay and angle of the user group. We compare our proposed method with the channel estimation method based on the tensor method. The simulation results show that compared with other methods, this method has obvious performance advantages in terms of estimation accuracy.

In future work we will optimize the clustering of user groups and the method of building dictionaries, as this will directly affect the accuracy of our channel estimation for user groups. In addition, we may also consider more diverse user group scenarios, such as dense groups of people in small areas or distant groups of people in large areas. A more optimal manifold approach will be considered as users are constantly on the move.

## REFERENCES

- [1] Y. Zhu, Z. Zhang, Z. Marzi, C. Nelson, U. Madhow, B. Y. Zhao, and H. Zheng, "Demystifying 60 GHz outdoor picocells," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 5–16.
- [2] 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, Feb. 2014.
- [3] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1444, Apr. 2013.
- [4] J. Zhong, G. Chen, J. Mao, S. Dang, and P. Xiao, "Iterative frequency domain equalization for MIMO-GFDM systems," *IEEE Access*, vol. 6, pp. 19386–19395, 2018.
- [5] H. Luo, Y. Zhang, W. Li, L. K. Huang, J. Cosmas, D. Li, and C. Maple, "Low latency parallel turbo decoding implementation for future terrestrial broadcasting systems," *IEEE Trans. Broadcast.*, vol. 64, no. 1, pp. 96–104, Mar. 2018.
- [6] A. Koochakzadeh and P. Pal, "Channel estimation for hybrid MIMO communication with (Non-) uniform linear arrays via tensor decomposition," in *Proc. IEEE 11st Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2020, pp. 1–5.
- [7] J. Du, S. Ye, Y. Chen, and Y. Guan, "Joint channel estimation techniques for multi-user massive MIMO relay networks," in *Proc. IEEE 20th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2020, pp. 29–33.
- [8] H. T. Dao and S. Kim, "Pilot power allocation for enhancing channel estimation quality in multi-cell multi-user massive MIMO systems," in *Proc. 25th Asia-Pacific Conf. Commun. (APCC)*, Nov. 2019, pp. 207–210.
- [9] X. Gao, L. Dai, S. Zhou, A. M. Sayeed, and L. Hanzo, "Wideband beamspace channel estimation for millimeter-wave MIMO systems relying on lens antenna arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 18, pp. 4809–4824, Sep. 2019.
- [10] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, 2016.
- [11] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.
- [12] M. Pesavento, A. B. Gershman, and M. Haardt, "Unitary root-MUSIC with a real-valued eigendecomposition: A theoretical and experimental performance study," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1306–1314, May 2000.
- [13] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.
- [14] D. Fan, F. Gao, G. Wang, Z. Zhong, and A. Nallanathan, "Angle domain signal processing aided channel estimation for indoor 60 GHz TDD/FDD massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1948–1961, Jun. 2017.
- [15] Z. Guo, X. Wang, and W. Heng, "Millimeter-wave channel estimation based on 2-D beamspace MUSIC method," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5384–5394, Aug. 2017.
- [16] R. Shafin, L. Liu, and J. Zhang, "DoA estimation and RMSE characterization for 3D massive-MIMO/FD-MIMO OFDM system," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 1–6.
- [17] Y. You, L. Zhang, and M. Liu, "IP aided OMP based channel estimation for millimeter wave massive MIMO communication," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.
- [18] Y. Shen and M. Z. Win, "Fundamental limits of wideband localization—Part I: A general framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4956–4980, Oct. 2010.
- [19] R. K. Martin, J. S. Velotta, and J. F. Raquet, "Bandwidth efficient cooperative TDOA computation for multicarrier signals of opportunity," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2311–2322, Jun. 2009.
- [20] J. Xiong and K. Jamieson, "Towards fine-grained radio-based indoor location," in *Proc. 12nd Workshop Mobile Comput. Syst. Appl. (HotMobile)*, 2012, p. 13.
- [21] D. C. Araújo, A. L. F. de Almeida, J. P. da Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42133–42147, 2019.
- [22] F. Wen and C. Liang, "Improved tensor-MODE based direction-of-arrival estimation for massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2182–2185, Dec. 2015.

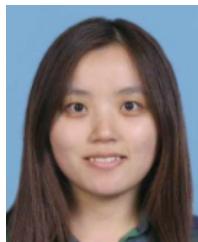
- [23] L. Cheng, Y.-C. Wu, J. Zhang, and L. Liu, "Subspace identification for DOA estimation in massive/full-dimension MIMO systems: Bad data mitigation and automatic source enumeration," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5897–5909, Nov. 2015.
- [24] L. Cheng, Y.-C. Wu, S. Ma, J. Zhang, and L. Liu, "Channel estimation in full-dimensional massive MIMO system using one training symbol," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Hokkaido, Japan, Jul. 2017, pp. 1–5.
- [25] Z. Dou, C. Li, C. Li, X. Gao, and L. Qi, "Tensor communication waveform design with semi-blind receiver in the MIMO system," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1727–1740, Feb. 2020.
- [26] C. Li, Z. Dou, G. Si, and M. Wang, "A novel digital-analog integrated communication system based on the PARATUCK-2 model," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–4.
- [27] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Structured tensor decomposition-based channel estimation for wideband millimeter wave MIMO," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 421–426.
- [28] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.
- [29] G. Favier and A. L. De Almeida, "Overview of constrained PARAFAC models," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, p. 142, 2014.
- [30] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.
- [31] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.
- [32] M. Ruble and I. Guvenc, "Multilinear singular value decomposition for millimeter wave channel parameter estimation," *IEEE Access*, vol. 8, pp. 75592–75606, 2020.
- [33] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-based channel estimation for millimeter wave MIMO-OFDM with dual-wideband effects," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4218–4232, Jul. 2020.
- [34] L. Cheng and Q. Shi, "Towards overfitting avoidance: Tuning-free tensor-aided multi-user channel estimation for 3D massive MIMO communications," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 832–846, Apr. 2021.
- [35] F. Wen, N. Garcia, J. Kulmer, K. Witrisal, and H. Wymeersch, "Tensor decomposition based beamspace ESPRIT for millimeter wave MIMO channel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–7.
- [36] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [37] L. D. Lathauwer, "Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, Sep. 2008.
- [38] T. G. Kolda, "Multilinear operators for higher-order decompositions," United States Dept. Energy, Washington, DC, USA, Tech. Rep., 2006.
- [39] M. R. Akdeniz, Y. Liu, M. K. Samimi, and S. Sun, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
- [40] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brisbane, Aust., Apr. 2015, pp. 2909–2913.
- [41] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [42] J. Palacios, N. Gonzalez-Prelcic, and J. Widmer, "Managing hardware impairments in hybrid millimeter wave MIMO systems: A dictionary learning-based approach," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 168–172.
- [43] M. Nazzal, O. Hasekioglu, A. R. Ekti, A. Gorcin, and H. Arslan, "Compressed spectrum sensing using sparse recovery convergence patterns through machine learning classification," in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2019, pp. 1–6.
- [44] K. You, W. Guo, T. Peng, Y. Liu, P. Zuo, and W. Wang, "Parametric sparse Bayesian dictionary learning for multiple sources localization with propagation parameters uncertainty," *IEEE Trans. Signal Process.*, vol. 68, pp. 4194–4209, 2020.
- [45] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vis. (ICCV)*, Sep./Oct. 2009, pp. 2272–2279.
- [46] P. Zhang, H. Wang, Y. Zhou, X. Sun, and H. Wang, "An improved three-dimensional auto-clustering algorithm for indoor millimeter-wave multipath propagation analysis," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Oct. 2017, pp. 1–6.
- [47] Y. Qiao, Z. Yang, T. Zhou, S. Lin, and D. Zhu, "A novel density-based performance evaluation index for multipath components clustering," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2020, pp. 499–503.
- [48] N. Czink, P. Cera, J. Salo, E. Bonek, J. Nuutila, and J. Ylitalo, "A framework for automatic clustering of parametric MIMO channel data including path powers," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2006, pp. 1–5.
- [49] D. Arthur and S. Vassilvitskii, *K-Means++: The Advantages of Careful Seeding*. Philadelphia, PA, USA: SIAM, 2006.
- [50] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.
- [51] D. Meng, G. Cao, W. Cao, and Z. He, "Supervised feature learning network based on the improved LLE for face recognition," in *Proc. Int. Conf. Audio, Lang. Image Process. (ICALIP)*, Shanghai, China, Jul. 2016, pp. 306–311.
- [52] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, Oct. 2017.
- [53] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5249–5257.
- [54] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, Mar. 2015.
- [55] J.-D. Lin, W.-H. Fang, and J.-T. Chen, "Constrained TST MUSIC for joint spatial-temporal channel parameter estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2003, pp. 193–196.
- [56] B. W. Bader and T. G. Kolda, "Algorithm 862: MATLAB tensor classes for fast algorithm prototyping," *ACM Trans. Math. Softw.*, vol. 32, no. 4, pp. 635–653, Dec. 2006.
- [57] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.



**XIAOPING ZHOU** received the Ph.D. degree in information and communication engineering from the University of Shanghai, Shanghai, China, in 2011. From 2011 to 2013, he was a Postdoctoral Fellow with the Communication Laboratory, Shanghai Jiao Tong University of Science and Technology, China. He is currently a Full Professor with the Shanghai Normal University of Information Science and Technology, Shanghai. His current research interests include mobile communication systems, image processing, parameter estimation, and electrostatic discharge.



**HAICHAO LIU** was born in Hebei, China, in 1996. He is currently pursuing the M.S. degree from Shanghai Normal University (SHNU), Shanghai, China. His current research interests include massive MIMO, millimeter-wave communications, and tensor analysis.



**BIN WANG** received the Ph.D. degree from the Department of Automation, Shanghai Jiao Tong University, Shanghai, China. She is currently a Lecturer with Shanghai Normal University, China. Her research interests include computer vision, machine learning, image processing, and multimedia analysis.



**YANG WANG** received the Ph.D. degree in information and communication engineering from the Shanghai Institute of Microsystems and Information Technology, Chinese Academy of Sciences, in 2015. Her current research interests include electromagnetic inverse scattering for imaging and computational processing.

• • •



**JIFENG HUANG** received the B.S. degree in radio engineering from Zhengzhou University, Zhengzhou, Henan, China, in 1984, the M.S. degree in communication and electronic system from Xi'an Jiaotong University, Xi'an, Shaanxi, China, in 1989, and the Ph.D. degree in measurement and control technology and automation instrument from the East China University of Science and Technology, Shanghai, China, in 2006. From 1989 to 1999, he was a Teacher with the Zhengzhou University of Aeronautics, Zhengzhou. Since 1999, he has been a Professor with the College of Information, Mechanical and Electrical Engineering, Shanghai Normal University. His research interests include pattern recognition, machine learning, and automation instrument. He received the Award for Scientific and Technological Advancement from the Aviation Ministry of China.