WILEY | Hindawi

Research Article

Grouping-Based Channel Estimation and Tracking for Millimeter Wave Massive MIMO Systems

Rui Yin [ID],[1] Xin Zhou,[2] Wei Qi,[1] Celimuge Wu [ID],[3] and Yunlong Cai[2]

[1]*Zhejiang University City College, China*
[2]*Zhejiang University, China*
[3]*The University of Electro-Communications, Japan*

Correspondence should be addressed to Celimuge Wu; celimuge@uec.ac.jp

Although the millimeter wave (mmWave) massive multiple-input and multiple-output (MIMO) system can potentially boost the network capacity for future communications, the pilot overhead of the system in practice will greatly increase, which causes a significant decrease in system performance. In this paper, we propose a novel grouping-based channel estimation and tracking approach to reduce the pilot overhead and computational complexity while improving the estimation accuracy. Specifically, we design a low-complexity iterative channel estimation and tracking algorithm by fully exploiting the sparsity of mmWave massive MIMO channels, where the signal eigenvectors are estimated and tracked based on the received signals at the base station (BS). With the recovered signal eigenvectors, the celebrated multiple-signal classification (MUSIC) algorithm can be employed to estimate the direction of arrival (DoA) angles and the path amplitude for the user terminals (UTs). To improve the estimation accuracy and accelerate the tracking speed, we develop a closed-form solution for updating the step-size in the proposed iterative algorithm. Furthermore, a grouping method is proposed to reduce the number of sharing pilots in the scenario of multiple UTs to shorten the pilot overhead. The computational complexity of the proposed algorithm is analyzed. Simulation results are provided to verify the effectiveness of the proposed schemes in terms of the estimation accuracy, tracking speed, and overhead reduction.

1. Introduction

The fifth generation (5G) wireless communication system has been proposed to meet the increasing demands on spectrum efficiency (SE), energy efficiency (EE), and network efficiency (NE). Millimeter wave (mmWave) communications operating in the frequency band spanning from 30 to 300 GHz have received great attention for 5G and beyond 5G (B5G) systems, since they are readily combined with massive multiple-input multiple-output (MIMO) [1, 2] techniques for dramatically improving the system SE, EE, and latency performance [3, 4].

The channel state information (CSI) accuracy is very important for the design of mmWave massive MIMO systems, and it plays a key role to achieve high SE and EE [5, 6]. Particularly, obtaining accurate CSI is the premise to leverage the full benefits brought by the massive MIMO technique on the mmWave frequency band [7]. Conventionally, CSI is commonly obtained with the assistance of the pilot sequences. Attributed to the orthogonality of pilot sequences, the CSI estimation with high accuracy can be obtained. Nonetheless, there is a positive linear relationship between the pilot overhead and the number of antennas. Therefore, when hundreds of antennas are deployed at the BS, the pilot overhead increases linearly, which may deteriorate the system performance significantly [8] due to the pilot contamination.

To reduce the pilot overhead, proper allocation of pilot sequences is one of the main solutions. The authors in [9] have designed a pilot reuse algorithm for massive MIMO with in-phase and quadrature-phase imbalances (IQI). Moreover, a pilot reuse method has been proposed in [8], where the channel power distribution in angular domain is derived via the channel covariance matrix analysis.

Accordingly, a pilot reuse pattern has been designed to reduce pilot overhead. In [10, 11], a novel dynamic pilot allocation scheme has been proposed with low complexity. A fractional pilot reuse scheme based on threshold optimization has been developed in [12], where different pilots are allocated to the users classified based on the fluid model.

In fact, the pilot reuse scheme unavoidably affects the accuracy of CSI estimation, which motivates the blind CSI estimation methods developed to reduce the estimation error without using pilot sequences, such as the subspace method (SM) and monte carlo methods (MCMs). However, these conventional methods have extremely high computational complexity due to the matrix decomposition and are not suitable to implement in practice. To reduce the computational complexity, a spatial basis expansion model (SBEM) has been proposed to model the mmWave massive MIMO channel with a uniform linear array (ULA) of antennas in [13]. Then, a unified transmission strategy for the mmWave massive MIMO systems in the spatial domain has been proposed, which includes low complex CSI estimation and user scheduling. The same channel model has been extended to the uniform rectangular array- (URA-) based massive MIMO systems in [14]. By combining two-dimensional fast Fourier transformation (2D-FFT) and spatial spectrum analysis, the CSI can be estimated with low-complexity based on 2D-FFT operation in [15].

By exploiting the sparse characteristic of mmWave channel, an open-loop channel estimator based on orthogonal matching pursuit has been proposed for mmWave hybrid MIMO system [16]. Based on the channel reciprocity characteristics in TDD systems, the Arnoldi iteration has been used to estimate the left and right singular subspaces of the mmWave MIMO channel in [17]. To reduce the computational complexity, a sequential and adaptive subspace estimation (SASE) method has been proposed to estimate the column and row subspaces of the mmWave MIMO channel in [18].

Most of the aforementioned works cannot track the variation on the CSI, which needs to be reimplemented after the channel coherence time to catch up with the varying channel coefficients. Therefore, tracking and simultaneously estimating the CSI can further reduce the computational complexity to achieve accurate CSI. In [19], a modified unscented Kalman filter has been proposed to track the DoA angles with known channel power gain as *a priori* information. In [20], the predetermined user motion model has been exploited to predict the change in the CSI of massive MIMO systems. The assumptions in both works hinder their practical operation. A sparse complex-valued neural network has been applied to approximate the uplink-to-downlink mapping function in [21]. However, the overhead required for the uplink channel estimation is also intolerable. In [22], a principal subspace analysis- (PSA-) based iterative method has been applied to estimate and track the CSI. However, the computational complexity of this method is still very high due to the eigenvalue decomposition (EVD), and the step size of the iterative algorithm is not optimized.

To reduce the computational complexity and the pilot overhead, we propose a novel channel estimation and tracking algorithm for the mmWave massive MIMO systems with URA antennas at the BS. First, the channel sparsity of the mmWave frequency channel is fully exploited. Based on the principal component analysis (PCA) method, a weighted iterative method with the optimized step size is proposed to estimate and track the signal eigenvectors. Second, the obtained eigenvectors are used to estimate DoA angles and path amplitudes via the classic multiple-signal classification (MUSIC) algorithm. Then, a standard inner product is used to match the elevation DoA angles with the right azimuth DoA angles on the same path. Moreover, a user terminal (UT) grouping scheme is proposed to realize the training sequence reuse in the scenario of multiple UTs, which substantially reduces the number of required pilots and the pilot contamination. The main contributions of this work can be summarized as follows.

(1) Based on the PCA, a weighted iterative method is proposed to estimate and track the eigenvectors of the signal subspace, which are used to recover the CSI. Since the complex matrix decomposition operation is avoided, the computational complexity decreases significantly

(2) To accelerate the convergence speed for the proposed iterative scheme, we develop a closed-from solution for updating the step-size in the proposed iterative algorithm

(3) To reduce the pilot overhead, a UT grouping method based on the spatial separability is proposed to divide the UTs into different groups. The spatial separability implies that the UTs in the same group do not have overlapping paths so that they can share the same pilot sequence and transmit simultaneously without causing interference

This paper is organized as follows. In Section 2, a mmWave massive MIMO system is described. The singular value decomposition (SVD) method and the method to estimate the CSI based on the signal eigenvectors are introduced in Section 3. In Section 4, a novel iterative method is proposed. The simulation results are demonstrated in Section 6. Finally, conclusions are drawn in Section 7.

2. System Model

In this section, we describe the wireless channel model of the investigated mmWave massive MIMO system. As shown in Figure 1, the BS equipped with a URA of $M \times N$ antennas serves UTs, each of which is equipped with a single antenna. Moreover, the wavelength is denoted as λ, and the interantenna spacing of both elevation and horizontal directions is denoted as d_s which is set to be half of the λ.

We assume that the signal transmitted from each UT, $i \in \mathscr{I} \triangleq \{1, \cdots, U\}$ where U is the number of UTs, to the BS arrives along P_i clusters consisting of a number of independent resolvable paths, and each path has a corresponding elevation DoA angle and azimuth DoA angle. Moreover, the mean and angle spread (AS) of the
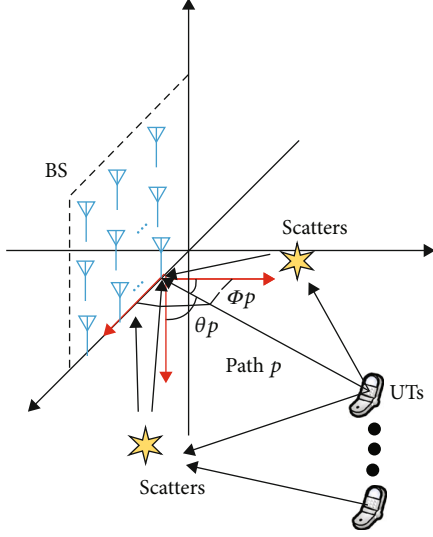
FIGURE 1: Massive MIMO system with URA of antennas at the BS.

elevation DoA angle of the p-th cluster are denoted by θ_p and $\Delta\theta_p$, respectively. Similarly, the mean and AS of the azimuth DoA angle of the p-th cluster are denoted by ϕ_p and $\Delta\phi_p$, respectively. When the signal propagates along the pth cluster, the channel can be written by integrating over the angular region as

$$\mathbf{H}_p = \int_{\theta_p - \Delta\theta_p}^{\theta_p + \Delta\theta_p} \int_{\phi_p - \Delta\phi_p}^{\phi_p + \Delta\phi_p} \beta(\theta, \phi)\mathbf{a}(\theta)\mathbf{b}^T(\theta, \phi)d\theta d\phi, \tag{1}$$

where $\mathbf{H}_p$ denotes the channel matrix corresponding to the pth cluster, $\beta(\theta, \phi) = |\alpha(\theta, \phi)| \exp\{-j\Phi(\theta, \phi)\}$, $\alpha(\theta, \phi)$ is the attenuation and $\Phi(\theta, \phi)$ is the initial phase, and $\mathbf{a}(\theta)$ and $\mathbf{b}^T(\theta, \phi)$ are the array manifold vectors (AMVs), which are written as

$$\mathbf{a}(\theta) = \left[1, e^{j\chi \cos(\theta)}, \cdots, e^{j\chi(M-1) \cos(\theta)}\right]^T, \tag{2}$$

$$\mathbf{b}(\theta, \phi)) = \left[1, e^{j\chi \sin(\theta) \cos(\phi)}, \cdots, e^{j\chi(N-1) \sin(\theta) \cos(\phi)}\right]^T, \tag{3}$$

where $\chi \triangleq 2\pi(d/\lambda)$.

Then, we define a steering matrix denoted as $\mathbf{A}(\theta, \phi) = \mathbf{a}(\theta)\mathbf{b}^T(\theta, \phi) \in \mathbb{C}^{M \times N}$ on the basis of AMVs, which can be written as

$$\begin{bmatrix} 1 & \cdots & e^{j\chi(N-1) \sin(\theta) \cos(\phi)} \\ e^{j\chi \cos(\theta)} & \cdots & e^{j\chi(\cos(\theta)+(N-1) \sin(\theta) \cos(\phi))} \\ \vdots & \vdots & \vdots \\ e^{j\chi(M-1) \cos(\theta)} & \cdots & e^{j\chi((M-1) \cos(\theta)+(N-1) \sin(\theta) \cos(\phi))} \end{bmatrix}. \tag{4}$$

Based on (4), the channel for the p-th cluster can be rewritten as

$$\mathbf{H}_p = \int_{\theta_p - \Delta\theta_p}^{\theta_p + \Delta\theta_p} \int_{\phi_p - \Delta\phi_p}^{\phi_p + \Delta\phi_p} \beta(\theta, \phi)\mathbf{A}(\theta, \phi)d\theta d\phi. \tag{5}$$

Based on (5), we can observe that the channel matrix, $\mathbf{H}_p$, mainly depends on $\beta(\theta, \phi)$ and $\mathbf{A}(\theta, \phi)$, where the former is directly related to the channel attenuation and initial phase, and the latter is determined by the DoA angles. Note that if $\mathbf{A}(\theta, \phi)$ and $\beta(\theta, \phi)$ are estimated with high accuracy, the CSI matrix would be reconstructed with high quality. To recover the CSI, there are six parameters to be estimated, i.e., θ_p, $\Delta\theta_p$, ϕ_p, $\Delta\phi_p$, $\alpha(\theta, \phi)$, and $\Phi(\theta, \phi)$.

Actually, (5) can be simplified when a mmWave frequency band is used in the massive MIMO system. From [2], $\Delta\theta_p$ and $\Delta\phi_p$ tend to be zero in the scenario where the massive MIMO system operates over the mmWave frequency band. Therefore, each cluster can be seen as an independent path, and each path is formed by the line-of-sight (LOS) and the first-order reflection components. Consequently, (5) can be rewritten as

$$\mathbf{H}_p = \beta\left(\theta_p, \phi_p\right)\mathbf{A}\left(\theta_p, \phi_p\right). \tag{6}$$

Then, the received signals at the BS is expressed as

$$\mathbf{Y}_k = \sum_{i=1}^{U} \sum_{p=1}^{P_i} \beta\left(\theta_{i,p}, \phi_{i,p}\right)\mathbf{A}\left(\theta_{i,p}, \phi_{i,p}\right)x_{i,k} + \mathbf{W} = \sum_{i=1}^{U} \sum_{p=1}^{P_i} \mathbf{H}_{i,p}x_{i,k} + \mathbf{W}, \tag{7}$$

where k is the index of the received signal snapshot, $x_{i,k}$ is the transmitted signal from UT i, $\mathbf{W}$ is the Additive White Gaussian Noise (AWGN) matrix, and $\mathbf{H}_{i,p} \in \mathbb{C}^{M \times N}$ is the channel matrix for UT i on the pth path. From (7), we can observe that to recover the channel matrix, only four parameters need to be estimated for each path p, which are $\theta_{i,p}$, $\phi_{i,p}$, $\alpha(\theta_{i,p}, \phi_{i,p})$, and $\Phi(\theta_{i,p}, \phi_{i,p})$. Therefore, by comparing with estimating entries in the original channel matrix, the computational complexity of recovering the parameters related to the paths is significantly reduced.

3. PCA-Based Eigenvectors Estimation and Tracking

In this section, a PCA-based iterative method to estimate and track the signal eigenvectors is proposed.

3.1. Channel Decomposition. Conventionally, SVD operation can be applied to estimate the eigenvectors related to the channel, which can be used to recover the channel. To explain the related notations and theories clearly, we assume that there is one UT in the system with P distinguishable paths first.

According to the SVD operation, the received signal, $\mathbf{Y}_k$, can be written as $\mathbf{Y}_k = \mathbf{SVD}^H$, where $\mathbf{S}$ is a unitary matrix $\mathbf{S} \in \mathbb{C}^{M \times M}$, $\mathbf{D}$ is a unitary matrix with $\mathbf{D} \in \mathbb{C}^{N \times N}$, the column

vectors of $\mathbf{S}$ and $\mathbf{D}$ are all referred to as singular vectors, and $\mathbf{V}$ is a diagonal matrix, $\mathbf{V} \in \mathbb{C}^{M \times N}$, which is named as singular value matrix and sorted in descending order.

Besides, when using the EVD to decompose the covariance matrices, $\mathbf{C}_L = \mathbb{E}\{\mathbf{Y}_k \mathbf{Y}_k^H\}$ and $\mathbf{C}_R = \mathbb{E}\{\mathbf{Y}_k^H \mathbf{Y}_k\}$, the column vectors of $\mathbf{S}$ and $\mathbf{D}$ are the eigenvectors of $\mathbf{C}_L$ and $\mathbf{C}_R$, respectively. Based on the array signal processing knowledge, the first P column vectors of $\mathbf{S}$ and $\mathbf{D}$ are mainly dependent on elevation and azimuth DoA angles of P resolvable paths, which are denoted as $\mathbf{S}_P$ and $\mathbf{D}_P$, respectively. Accordingly, the DoA angles and amplitudes of resolvable paths can be estimated after recovering the $\mathbf{S}_P$ and $\mathbf{D}_P$.

Based on above investigation, by applying the EVD operation, the covariance matrix, $\mathbf{C}_L$, can be written as

$$\mathbf{C}_L = \mathbf{S}_s \mathbf{V}_s \mathbf{S}_P^H + \mathbf{S}_z \mathbf{V}_{z_L} \mathbf{S}_z^H, \tag{8}$$

where matrix $\mathbf{S}_P \in \mathbb{C}^{M \times P}$, and the column vectors of $\mathbf{S}_P$ are eigenvectors, $\mathbf{V}_s$ is a diagonal matrix and $\mathbf{V}_s \in \mathbb{C}^{P \times P}$ with entries from λ_1 to λ_P, $\mathbf{S}_Z$ is a matrix $\in \mathbb{C}^{M \times (M-P)}$, and the column vectors of $\mathbf{S}_Z$ are eigenvectors, and $\mathbf{V}_{z_L}$ is a diagonal matrix $\mathbf{V}_{z_L} \in \mathbb{C}^{(M-P) \times (M-P)}$ with entries of eigenvalues from λ_{P+1} to λ_M.

Similarly, applying EVD operation to $\mathbf{C}_R$, we can obtain

$$\mathbf{C}_R = \mathbf{D}_P \mathbf{V}_d \mathbf{D}_P^H + \mathbf{D}_z \mathbf{V}_{z_R} \mathbf{D}_z^H, \tag{9}$$

where $\mathbf{D}_P \in \mathbb{C}^{N \times P}$, and the column vectors in $\mathbf{D}_P$ are eigenvectors, $\mathbf{V}_d$ is a diagonal matrix, $\mathbf{V}_d \in \mathbb{C}^{P \times P}$, with entries from γ_1 to γ_P, $\mathbf{D}_Z \in \mathbb{C}^{M \times (N-P)}$, and the column vectors of $\mathbf{S}_Z$ are eigenvectors, $\mathbf{V}_{z_R}$ is a diagonal matrix, and $\mathbf{V}_{z_R} \in \mathbb{C}^{(N-P) \times (N-P)}$, with entries of eigenvalues from γ_{P+1} to γ_M.

Even though $\mathbf{S}_P$ and $\mathbf{D}_P$ can be recovered via the SVD or EVD operation as analyzed above, each of them has very high computational complexity, which is $\mathbb{O}(M^3)$ and $\mathbb{O}(M^3 + N^3)$, respectively. Therefore, they are not suitable for channel estimation in practice. Moreover, it cannot trace the change on the channel matrix caused by the mobility of UTs.

3.2. Adaptive PCA-Based Iterative Method.
In order to estimate and track the eigenvectors with low computational complexity, an adaptive PCA-based iterative method is proposed in this subsection. First, we take the estimation and tracking of $\mathbf{S}_P$ for instance. Define a matrix $\mathbf{T}_1 = \mathbf{S}_P \mathbf{U} \in \mathbb{C}^{M \times P}$, where $\mathbf{U}$ is a unitary matrix with $P \times P$ dimension. The estimation on $\mathbf{T}_1$ can be formulated as the following optimization problem

$$\min_{\mathbf{T}_1} J(\mathbf{T}_1) = \mathbb{E}\left\{ tr\left(\left\| \mathbf{Y}_k - \mathbf{T}_1 \mathbf{T}_1^H \mathbf{Y}_k \right\|^2 \right) \right\}$$
$$= tr(\mathbf{C}_L) - 2tr(\mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1) + tr(\mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1 \mathbf{T}_1^H \mathbf{T}_1), \tag{10}$$

which is to minimize the mean square error (MSE) of the estimation. A simple iterative gradient descent method can

be used to solve the optimization problem in (10). The gradient of $J(\mathbf{T}_1)$ with respect to $\mathbf{T}_1$ is updated by

$$\mathbf{T}_{1,j} = \mathbf{T}_{1,j-1} - \eta \left. \frac{\partial J(\mathbf{T}_1)}{\partial \mathbf{T}_1} \right|_{\mathbf{T}_1 = \mathbf{T}_{1,j-1}}, \tag{11}$$

where j denotes the iteration index, $\eta > 0$ is a step size, and $(\partial J(\mathbf{T}_1))/(\partial \mathbf{T}_1)$ is to represent the gradient of $J(\mathbf{T}_1)$ with respect to $\mathbf{T}_1$. To obtain the partial derivative of $J(\mathbf{T}_1)$ with respect to $\mathbf{T}_1$, we first calculate the differential of $J(\mathbf{T}_1)$ with respect to $\mathbf{T}_1$ via the chain rule, which is given by

$$\begin{aligned} dJ(\mathbf{T}_1) = {} & -2tr\left(\mathbf{T}_1^H\left(\mathbf{C}_L^H + \mathbf{C}_L\right)d\mathbf{T}_1\right) + tr\left(d\mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1 \mathbf{T}_1^H \mathbf{T}_1\right) \\ & + tr\left(\mathbf{T}_1^H \mathbf{C}_L d\mathbf{T}_1 \mathbf{T}_1^H \mathbf{T}_1\right) + tr\left(\mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1 d\mathbf{T}_1^H \mathbf{T}_1\right) \\ & + tr\left(\mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1 \mathbf{T}_1^H d\mathbf{T}_1\right) = -4tr\left(\mathbf{T}_1^H \mathbf{C}_L d\mathbf{T}_1\right) \\ & + 2tr\left(\mathbf{T}_1^H \mathbf{T}_1 \mathbf{T}_1^H \mathbf{C}_L d\mathbf{T}_1\right) + 2tr\left(\mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1 \mathbf{T}_1^H d\mathbf{T}_1\right). \end{aligned} \tag{12}$$

Based on the matrix analysis, the gradient of $J(\mathbf{T}_1)$ is given by

$$\frac{\partial J(\mathbf{T}_1)}{\partial \mathbf{T}_1} = -2\left(2\mathbf{C}_L - \mathbf{C}_L \mathbf{T}_1 \mathbf{T}_1^H - \mathbf{T}_1 \mathbf{T}_1^H \mathbf{C}_L\right)\mathbf{T}_1. \tag{13}$$

After substituting (13) into (11), the update rule on $\mathbf{T}_1$ can be rewritten as

$$\mathbf{T}_{1,j} = \mathbf{T}_{1,j-1} + 2\eta \left(2\mathbf{C}_L - \mathbf{C}_L \mathbf{T}_{1,j-1} \mathbf{T}_{1,j-1}^H - \mathbf{T}_{1,j-1} \mathbf{T}_{1,j-1}^H \mathbf{C}_L\right)\mathbf{T}_{1,j-1}. \tag{14}$$

With Equation (14), $\mathbf{T}_1$ will converge to the principal subspace, $\mathbf{S}_P \mathbf{U}$, which is not equal to $\mathbf{S}_P$ due to the unitary matrix $\mathbf{U}$.

3.3. Optimization Problem.
To estimate the corresponding eigenvectors, we reconstruct the optimization problem based on the weighted method proposed in [23]. The new optimization problem is formulated as

$$\min_{\mathbf{T}_1} J(\mathbf{T}_1) = tr(\mathbf{C}_L) - 2tr(\boldsymbol{\Omega} \mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1) + tr(\boldsymbol{\Omega} \mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1 \mathbf{T}_1^H \mathbf{T}_1), \tag{15}$$

where $\boldsymbol{\Omega}$ is a diagonal matrix whose diagonal item is $w_{i,i}$. If the diagonal items of $\boldsymbol{\Omega}$ are set with different and positive values and $w_{1,1} > w_{2,2} > \cdots > w_{P,P}$, the estimation of $\mathbf{T}_1$ will converge to $\mathbf{S}_P$ [23]. Based on (15), the gradient of $J(\mathbf{T}_1)$ can be rewritten as

$$\begin{aligned} \frac{\partial J(\mathbf{T}_1)}{\partial \mathbf{T}_1} = {} & -\left(4\mathbf{C}_L \mathbf{T}_1 \boldsymbol{\Omega} - \mathbf{C}_L \mathbf{T}_1 \boldsymbol{\Omega} \mathbf{T}_1^H \mathbf{T}_1 - \mathbf{T}_1 \mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1 \boldsymbol{\Omega} \right. \\ & \left. - \mathbf{C}_L \mathbf{T}_1 \mathbf{T}_1^H \mathbf{T}_1 \boldsymbol{\Omega} - \mathbf{T}_1 \boldsymbol{\Omega} \mathbf{T}_1^H \mathbf{C}_L \mathbf{T}_1\right). \end{aligned} \tag{16}$$

With (16), matrix $\mathbf{T}_1$ in (14) is updated by

$$\begin{aligned}
\mathbf{T}_{1,j} = \mathbf{T}_{1,j-1} + \eta\Big(&4\mathbf{C}_L\mathbf{T}_{1,k-1}\mathbf{\Omega} - \mathbf{C}_L\mathbf{T}_{1,j-1}\mathbf{\Omega}\mathbf{T}_1^H\mathbf{T}_{1,j-1} \\
&- \mathbf{T}_{1,j-1}\mathbf{T}_{1,j-1}^H\mathbf{C}_L\mathbf{T}_{1,j-1}\mathbf{\Omega} - \mathbf{C}_L\mathbf{T}_{1,j-1}\mathbf{T}_{1,j-1}^H\mathbf{T}_{1,j-1}\mathbf{\Omega} \\
&- \mathbf{T}_{1,j-1}\mathbf{\Omega}\mathbf{T}_{1,j-1}^H\mathbf{C}_L\mathbf{T}_{1,j-1}\Big).
\end{aligned}$$

(17)

It has been proven in [23] that $\mathbf{T}_1$ converges to $\mathbf{S}_P$ when j tends to infinity. From (17), we can observe that the covariance matrix $\mathbf{C}_L$ is essential to update $\mathbf{T}_1$, but the accurate $\mathbf{C}_L$ is difficult to be obtained due to $\mathbf{C}_L$ is the expectation of $\mathbf{Y}_k\mathbf{Y}_k^H$. To obtain $\mathbf{C}_L$, an approximated method has been proposed in [24], which uses a sampling covariance matrix $\widehat{\mathbf{C}}_{L,j}$. The $\widehat{\mathbf{C}}_{L,j}$ is given by

$$\widehat{\mathbf{C}}_{L,j} = \frac{1}{j}\sum_{k=0}^{j}\tau^{j-i}\mathbf{Y}_k\mathbf{Y}_k^H = \tau\frac{j-1}{j}\widehat{\mathbf{C}}_{L,j-1} + \frac{1}{j}\mathbf{Y}_j\mathbf{Y}_j^H,$$

(18)

where $\tau \in [0,1]$ is the forgetting factor, the variable k is used to indicate the kth time slot, and j is to indicate the current time slot. The initial matrix $\widehat{\mathbf{C}}_{L,1}$ is set to be equal to $\mathbf{Y}_1\mathbf{Y}_1^H$. Moreover, it is noteworthy that the parameter τ can be adjusted to trace channel variance. If the channel is modeled as a stationary process, (18) can be regarded as the calculation on the average value of $\mathbf{Y}_k\mathbf{Y}_k^H$, and τ is equal to 1. However, the τ is taken in the range of (0,1) due to the massive MIMO channel is not stationary in practice.

With (18), the updating rule for $\mathbf{T}_1$ can be rewritten as

$$\begin{aligned}
\mathbf{T}_{1,k} = \mathbf{T}_{1,k-1} + \eta\Big(&4\widehat{\mathbf{C}}_{1,k-1}\mathbf{T}_{1,k-1}\mathbf{\Omega} - \widehat{\mathbf{C}}_{L,k-1}\mathbf{T}_{1,k-1}\mathbf{\Omega}\mathbf{T}_1^H\mathbf{T}_{1,k-1} \\
&- \mathbf{T}_{1,k-1}\mathbf{T}_{1,k-1}^H\widehat{\mathbf{C}}_{1,k-1}\mathbf{T}_{1,k-1}\mathbf{\Omega} - \widehat{\mathbf{C}}_{1,k-1}\mathbf{T}_{1,k-1}\mathbf{T}_{1,k-1}^H\mathbf{T}_{1,k-1}\mathbf{\Omega} \\
&- \mathbf{T}_{1,k-1}\mathbf{\Omega}\mathbf{T}_{1,k-1}^H\widehat{\mathbf{C}}_{1,k}\mathbf{T}_{1,k-1}\Big).
\end{aligned}$$

(19)

It is noteworthy that the step size, η, is essential to obtain $\mathbf{T}_1$ with high accuracy and convergence speed. However, deriving an appropriate step size η is nontrivial. On the one hand, if η is small, the convergence speed will be slow. On the other hand, the accuracy will be impaired if η is high. Therefore, in the next subsection, the optimal η is derived to accelerate the convergence speed while keeping the accuracy.

3.4. Optimal Step Size. To find the optimal step size η_{opt}, we consider the objection function $J(\mathbf{T}_1)$ at the $(k+1)$th iteration step which is given by

$$\begin{aligned}
J(\mathbf{T}_{1,k+1}) = tr(\mathbf{R}_{1,k+1}) &- 2tr(\mathbf{\Omega}\mathbf{T}_{1,k+1}^H\mathbf{R}_{1,k+1}\mathbf{T}_{1,k+1}) \\
&+ tr(\mathbf{\Omega}\mathbf{T}_{1,k+1}^H\mathbf{R}_{1,k+1}\mathbf{T}_{1,k+1}\mathbf{T}_{1,k+1}^H\mathbf{T}_{1,k+1}).
\end{aligned}$$

(20)

To simplify the equation, we replace $(\partial J(\mathbf{T}_1))/(\partial\mathbf{T}_1)$ with ∇. Then, combining (11) with (20), (20) can be rewritten as

$$\begin{aligned}
J(\mathbf{T}_{1,k+1},\eta) = tr(\widehat{\mathbf{C}}_{L,k+1}) &- 2tr\big[\mathbf{\Omega}(\mathbf{T}_{1,k} - \eta\nabla)^H\widehat{\mathbf{C}}_{L,k+1}(\mathbf{T}_{1,k} - \eta\nabla)\big] \\
&+ tr\big[\mathbf{\Omega}(\mathbf{T}_{1,k} - \eta\nabla)^H\widehat{\mathbf{C}}_{L,k+1}(\mathbf{T}_{1,k} - \eta\nabla)(\mathbf{T}_{1,k} - \eta\nabla)^H \\
&\cdot (\mathbf{T}_{1,k} - \eta\nabla)\big].
\end{aligned}$$

(21)

Since $\mathbf{T}_{1,k+1}^H\mathbf{T}_{1,k+1} = \mathbf{U}^H\mathbf{S}_P^H\mathbf{S}_P\mathbf{U} = \mathbf{I}$, (21) can be written as

$$J(\mathbf{T}_{1,k+1},\eta) = tr(\widehat{\mathbf{C}}_{L,k+1}) - tr\big[\mathbf{\Omega}(\mathbf{T}_{1,k} - \eta\nabla)^H\widehat{\mathbf{C}}_{L,k+1}(\mathbf{T}_{1,k} - \eta\nabla)\big].$$

(22)

Therefore, $J(\mathbf{T}_{1,k+1})$ has a global maximum with the optimal step size, η_{opt}. Based on (22), we can differentiate $J(\mathbf{T}_{1,k+1},\eta)$ with respect to η, which is given by

$$\begin{aligned}
\frac{\partial J(\mathbf{T}_{1,k+1},\eta)}{\partial\eta} &= \frac{\partial}{\partial\eta}\Big\{-2tr\big[\mathbf{\Omega}(\mathbf{T}_{1,k} - \eta\nabla)^H\widehat{\mathbf{C}}_{L,k+1}(\mathbf{T}_{1,k} - \eta\nabla)\big] \\
&\quad + tr\big[\mathbf{\Omega}(\mathbf{T}_{1,k} - \eta\nabla)^H\widehat{\mathbf{C}}_{L,k+1}(\mathbf{T}_{1,k} - \eta\nabla) \\
&\quad \cdot (\mathbf{T}_{1,k} - \eta\nabla)^H(\mathbf{T}_{1,k} - \eta\nabla)\big]\Big\} \\
&= -2tr\big[\mathbf{\Omega}(-\nabla)^H\widehat{\mathbf{C}}_{1,k+1}(\mathbf{T}_{1,k} - \eta\nabla)\big] \\
&\quad - 2tr\big[\mathbf{\Omega}(\mathbf{T}_{1,k} - \nabla)^H\widehat{\mathbf{C}}_{L,k+1}(-\eta\nabla)\big] \\
&= 2tr\big[\mathbf{\Omega}\nabla^H\widehat{\mathbf{C}}_{L,k+1}\mathbf{T}_{1,k}\big] - 2\eta tr\big[\mathbf{\Omega}\nabla^H\widehat{\mathbf{C}}_{L,k+1}\nabla\big] \\
&\quad + 2tr\big[\mathbf{\Omega}\mathbf{T}_{1,k}^H\widehat{\mathbf{C}}_{L,k+1}\nabla\big] - tr\big[\mathbf{\Omega}\nabla^H\widehat{\mathbf{C}}_{L,k+1}\nabla\big].
\end{aligned}$$

(23)

When $(\partial J(\mathbf{T}_{1,k+1}))/(\partial\eta) = 0$, $J(\mathbf{T}_{1,k+1})$ has a global maximum. As a result, the optimal step size, η_{opt}, is the solution of $(\partial J(\mathbf{T}_{1,k+1}))/(\partial\eta) = 0$, which is given by

$$\eta_{opt} = \frac{tr\big[\mathbf{\Omega}\nabla^H\widehat{\mathbf{C}}_{L,k+1}\mathbf{T}_{1,k}\big] + tr\big[\mathbf{\Omega}\mathbf{T}_{1,k}^H\widehat{\mathbf{C}}_{L,k+1}\nabla\big]}{2tr\big[\mathbf{\Omega}\nabla^H\widehat{\mathbf{C}}_{L,k+1}\nabla\big]}.$$

(24)

Based on (24), we can achieve the optimal updating rule to estimate and track the signal eigenvector matrix, $\mathbf{S}_P$. In the same way, we can estimate and track the $\mathbf{D}_P$.

4. Estimation on the Channel Parameters

After estimating $\mathbf{S}_P$ and $\mathbf{D}_P$ in Section 3, the DoA angles of the resolvable paths on both elevation and azimuth directions can be recovered based on the classic MUSIC algorithm. Define u_p and v_p as

$$u_p = \chi\cos\theta_p,$$

(25)

$$v_p = \chi\sin\theta_p\cos\phi_p.$$

(26)

With these two variables, the steering matrix, $\mathbf{A}(\theta,\phi) = \mathbf{a}(\theta)\mathbf{b}^T(\theta,\phi)$, can be rewritten as $A(u_p,v_p) = \mathbf{a}(u_p)\mathbf{b}^T(v_p)$. Then, the MUSIC method is used to estimate u_p and v_p. Then, the DoA angles can be recovered according to (25) and (26).

4.1. Estimation on u_p and v_p. Multiplying $\mathbf{C}_L = \mathbb{E}\{\mathbf{Y}_k\mathbf{Y}_k^H\}$ and (8) with $\mathbf{S}_z$, we obtain

$$\mathbf{C}_L\mathbf{S}_z = \sum_{p=1}^{P} \mathbf{H}_p\psi_i\mathbf{H}_p^H\mathbf{S}_z + \sigma^2\mathbf{S}_z, \tag{27}$$

$$\mathbf{C}_L\mathbf{S}_z = \mathbf{S}_P\mathbf{\Lambda}_s\mathbf{S}_P^H\mathbf{S}_z + \mathbf{S}_z\mathbf{\Lambda}_{z_1}\mathbf{S}_z^H\mathbf{S}_z = \sigma^2\mathbf{S}_z. \tag{28}$$

where $\psi_i = \mathbb{E}\{x_ix_i^H\}$.

Then, subtracting (27) to (28), there is

$$\mathbf{H}_p^H\mathbf{S}_z = 0. \tag{29}$$

Since $\mathbf{h}_p = \beta(\theta_p, \phi_p)\mathbf{A}(u_p, v_p)$, we can derive the following equation from (29)

$$\mathbf{A}(u_p, v_p)^H\mathbf{S}_z = \mathbf{0}. \tag{30}$$

Based on matrix manipulation, from (30), we can obtain

$$\mathbf{a}^H(u_p)\mathbf{s}_{z,k} = \sum_{m=0}^{M-1} e^{jmu_p}s_{z,k,m} = 0, \tag{31}$$

where $\mathbf{s}_{z,k}$ is the kth column of matrix $\mathbf{S}_z$ and $s_{z,k,m}$ is the mth entry of $\mathbf{s}_{z,k}$.

Based on (31), the pseudospectrum with respect to u can be derived as

$$f(u) = \frac{1}{\sum_{k=P+1}^{M}\left|\mathbf{a}(u_p)^H\mathbf{s}_{z,k}\right|^2} = \frac{1}{\mathbf{p}^H(\zeta)\mathbf{S}_z\mathbf{S}_z^H\mathbf{p}(\zeta)}, \tag{32}$$

where $\mathbf{p}(\zeta) = [1, \zeta, \cdots, \zeta^{(M-1)}]^T$. According to (32), if $\zeta = e^{ju}$, the roots of the denominator would locate on the unit circle and be the estimation on u since the steering vector, $\mathbf{a}(u_p)$, $1 \le p \le P$, is orthogonal to the noise subspace, $\mathbf{S}_z$. Moreover, $\mathbf{S}_z\mathbf{S}_z^H$ can be derived based on the estimation on $\mathbf{S}_P$ in Section 2, which is given by

$$\mathbf{S}_z\mathbf{S}_z^H = \mathbf{I} - \mathbf{S}_P\mathbf{S}_P^H. \tag{33}$$

Then, inserting (33) into (32), u_p, $\forall 1 \le i \le P$, can be estimated by finding the peaks of the pseudospectrum in (32). By using the same method, we can estimate v_p, $\forall 1 \le p \le P$, with $\mathbf{D}_P$.

4.2. Matching on u_p and v_p. After estimating u_p and v_p in the above subsection, it needs to match u_p with the right v_p on the same path since each path has a corresponding pair of elevation and azimuth DoA angles. To do so, a standard inner product is defined to match u_p with v_p, which is given by

$$P_r = <\mathbf{J}, \mathbf{K}> = tr(\mathbf{J}^H\mathbf{K}), \tag{34}$$

where $\mathbf{J}$ and $\mathbf{K}$ are two matrices. The inner product is to calculate the energy of the received signals on certain direction.

Accordingly, with the estimation of u_p and v_p on the above subsection, the inner products of received signals matrix $\mathbf{Y}$ with the steering matrix $\mathbf{A}(u_i, v_j)$ is given by

$$\begin{aligned}P_r &= <\mathbf{A}(u_i, v_j), \mathbf{Y}> = tr(\mathbf{A}^H(u_i, v_j)\mathbf{Y}) = tr(\mathbf{b}(v_j)^*\mathbf{a}^H(u_i)\mathbf{Y}) \\ &= \mathbf{a}^H(u_i)\mathbf{Y}\mathbf{b}(v_j)^*,\end{aligned} \tag{35}$$

where $i, j = 1, 2, 3 \cdots P$. P_r will be the maximum value when $i = j$. The received signals from P paths make the most contribution to the received energy at the BS. Then, we can match u_p with v_p by searching the maximum value of P_r. After that, based on (25) and (26), the DoA angles of each path can be estimated by the following formulas

$$\theta_p = \arccos\left(\frac{u_p}{\chi}\right), \quad \forall 1 \le p \le P, \tag{36}$$

$$\phi_p = \arccos\left(\frac{v_p}{\chi\sqrt{1 - (u_p/\chi)^2}}\right), \quad \forall 1 \le p \le P. \tag{37}$$

4.3. Estimation on $\beta(\theta, \phi)$. Based on (7) and SVD operation, the received signals $\mathbf{Y}_k$ is rewritten as

$$\mathbf{Y}_k = \mathbf{S\Sigma D}^H = \sum_{p=1}^{P} \beta(\theta_p, \phi_p)\mathbf{A}(u_p, v_p)x_k + \mathbf{W} = \sum_{p=1}^{P} \rho_p\mathbf{s}_p\mathbf{d}_p^H. \tag{38}$$

where ρ_p is the pth eigenvalue of $\mathbf{Y}_k$; $\mathbf{s}_p$ and $\mathbf{d}_p$ are the pth eigenvectors of $\mathbf{S}_P$ and $\mathbf{D}_P$, respectively. Then, we set $\mathbf{G}_p = \mathbf{s}_p\mathbf{d}_p^H$ and construct an orthonormal space as follows

$$<\mathbf{G}_i, \mathbf{G}_j> = tr(\mathbf{G}_i^H\mathbf{G}_j) = \begin{cases} 0, & \text{if } i \ne j, \\ 1, & \text{if } i = j. \end{cases} \tag{39}$$

According to (39), the pth eigenvalue, ρ_p, can be written as

$$\begin{aligned}\rho_p &= <\mathbf{G}_p, \mathbf{Y}_k> = <\mathbf{G}_p, \sum_{p=1}^{P}\beta(\theta_p, \phi_p)\mathbf{A}(u_p, v_p)x + \mathbf{W}> \\ &= \sum_{p=1}^{P} \beta(\theta_p, \phi_p)\mathbf{s}_p^H\mathbf{A}(u_p, v_p)\mathbf{d}_px + \mathbf{s}_p^H\mathbf{W}\mathbf{d}_p \\ &= \sum_{p=1}^{P} \beta(\theta_p, \phi_p)\mathbf{s}_p^H\mathbf{A}(u_p, v_p)\mathbf{d}_px + \sigma^2.\end{aligned}$$

$$\tag{40}$$

1: Receive the kth signal snapshot in the kth time slot, $\mathbf{Y}_k$. Then, the BS updates the covariance matrix, $\widehat{\mathbf{C}}_{L,k}$, based on (18);
2: Estimate and track the eigenvector matrix of the received signal, $\mathbf{S}_P$, based on the PCA algorithm in (19). Recover $\mathbf{D}_P$ in the same way;
3: Estimate u_p and v_p, $1 \le i \le P$, based on (32);
4: Pair u_p, $1 \le p \le P$, with v_p, $1 \le p \le P$, based on (35);
5: Estimate θ_p, $1 \le p \le P$ and ϕ_p, $1 \le p \le P$ based on (36) and (37);
6: Estimate the amplitude corresponding to each path based on (43);
7: Recover the channel based on (6);

ALGORITHM 1: Channel estimation and tracking scheme

Based on (40), define

$$
\mathbf{O} = \det\left(\begin{bmatrix} \mathbf{s}_1^H \mathbf{A}(u_1,v_1)\mathbf{d}_1 & \mathbf{s}_1^H \mathbf{A}(u_2,v_2)\mathbf{d}_1 & \cdots & \mathbf{s}_1^H \mathbf{A}(u_P,v_P)\mathbf{d}_1 \\ \mathbf{s}_2^H \mathbf{A}(u_1,v_1)\mathbf{d}_2 & \mathbf{s}_2^H \mathbf{A}(u_2,v_2)\mathbf{d}_2 & \cdots & \mathbf{s}_2^H \mathbf{A}(u_P,v_P)\mathbf{d}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_P^H \mathbf{A}(u_1,v_1)\mathbf{d}_P & \mathbf{s}_P^H \mathbf{A}(u_2,v_2)\mathbf{d}_P & \cdots & \mathbf{s}_P^H \mathbf{A}(u_P,v_P)\mathbf{d}_P \end{bmatrix} \right),
$$

$$(41)$$

$$
\mathbf{O}_p = \det\left(\begin{bmatrix} \mathbf{s}_1^H \mathbf{A}(u_1,v_1)\mathbf{d}_1 & \cdots & \rho_1 - \sigma & \cdots & \mathbf{s}_1^H \mathbf{A}(u_P,v_P)\mathbf{d}_1 \\ \mathbf{s}_2^H \mathbf{A}(u_1,v_1)\mathbf{d}_2 & \cdots & \rho_2 - \sigma & \cdots & \mathbf{s}_2^H \mathbf{A}_P(u_P,v_P)\mathbf{d}_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_P^H \mathbf{A}(u_1,v_1)\mathbf{d}_P & \cdots & \rho_P - \sigma & \cdots & \mathbf{s}_P^H \mathbf{A}(u_P,v_P)\mathbf{d}_P \end{bmatrix} \right).
$$

$$(42)$$

Based on the Cramer's Rule, the amplitude of path i can be calculated as

$$
\beta\left(\theta_p, \phi_p\right) = \frac{\mathbf{O}_p}{\mathbf{O}}, \quad 1 \le p \le P. \tag{43}
$$

Then, the complete algorithm flow of the proposed scheme is given in Algorithm 1.

Algorithm 1: Channel estimation and tracking scheme1: Receive the kth signal snapshot in the kth time slot, $\mathbf{Y}_k$. Then, the BS updates the covariance matrix, $\widehat{\mathbf{C}}_{L,k}$, based on (18);

2: Estimate and track the eigenvector matrix of the received signal, $\mathbf{S}_P$, based on the PCA algorithm in (19). Recover $\mathbf{D}_P$ in the same way;

3: Estimate u_p and v_p, $1 \le i \le P$, based on (32);

4: Pair u_p, $1 \le p \le P$, with v_p, $1 \le p \le P$, based on (35);

5: Estimate θ_p, $1 \le p \le P$ and ϕ_p, $1 \le p \le P$ based on (36) and (37);

6: Estimate the amplitude corresponding to each path based on (43);

7: Recover the channel based on (6);

5. Extension for Multiple UTs

In this section, the proposed method is extended to the multiple UT scenario. When the UTs have the same paths, their signals cannot be distinguished at the BS. To solve this problem and reduce the pilot overhead, a user grouping scheme is proposed in this section.

5.1. Initial DoA Angle Estimation. The initial DoA angles of different UTs need to be known at the BS before dividing them into different groups. To reach this goal, the training process is necessary. Assume that there are I orthogonal training sequences, denoted as $\{\mathbf{r}_1, \cdots, \mathbf{r}_i, \cdots, \mathbf{r}_I\}$, where $\mathbf{r}_i$ is an L-dimensional vector with $\|\mathbf{r}_i\|^2 = L$. During the training stage, each UT is assigned to a training sequence, and the received signals at BS can be expressed as

$$
\bar{\mathbf{Y}} = \sum_{i \in \mathscr{U}} \bar{\mathbf{H}}_i \mathbf{r}_i + \bar{\mathbf{W}}, \tag{44}
$$

where $\bar{\mathbf{Y}} \in \mathbb{C}^{M \times N \times L}$, and $\bar{\mathbf{H}}_i \in \mathbb{C}^{M \times N \times 1}$ is the uplink channel between the ith UT and the BS; $\mathscr{U}$ is the set of UTs. $\bar{\mathbf{W}} \in \mathbb{C}^{M \times N \times L}$ is the additive white Gaussian noise. The orthogonality of the training sequences can be used to distinguish the received signals from different UTs. First, multiply (44) with $\mathbf{r}_i^H$ to get

$$
\bar{\mathbf{Y}}_i = \frac{1}{L}\bar{\mathbf{Y}}\mathbf{r}_i^H = \frac{1}{L}\left(\sum_{j \in \mathscr{U}} \bar{\mathbf{H}}_j \mathbf{r}_j + \bar{\mathbf{W}} \right)\mathbf{r}_i^H = \bar{\mathbf{H}}_i + \frac{1}{L}\bar{\mathbf{W}}\mathbf{r}_i^H. \tag{45}
$$

Based on (45), the transmitted signal from different UTs can be separated. With the detached signal $\bar{\mathbf{Y}}_i$, the proposed channel estimation method can be used to estimate the initial DoA angles associated with the ith UT. However, the number of UTs may be more than the number of orthogonal training sequences, i.e., $U > I$. To solve this problem, the pilot sequences can be reused in a time division multiple access (TDMA) manner. UTs are divided into $\lceil U/I \rceil$ groups, and the UTs in each group can reuse the same I orthogonal pilot sequences in different time slots. Accordingly, the BS can estimate the initial DoA angles of all UTs.

5.2. Grouping Scheme. After the initial DoA angle estimation, the UTs can be partitioned into different groups, and the UTs in the same group can use the same training sequence and transmit simultaneously.

1: Initialization: divide UTs into $\lceil U/I \rceil$ groups. Different UTs in the same group are allocated with different pilot sequences;
2: UTs in the same group transmit the pilots in the same time slot. The BS distinguishes the received signals from different UTs;
3: The BS estimates the initial DoA angles of different UTs;
4: If the minimum difference on the DoA angles between any two UTs is bigger than a threshold denoted as λ, they are put into the same group;
5: The UTs in the same group will use the same pilot sequence.
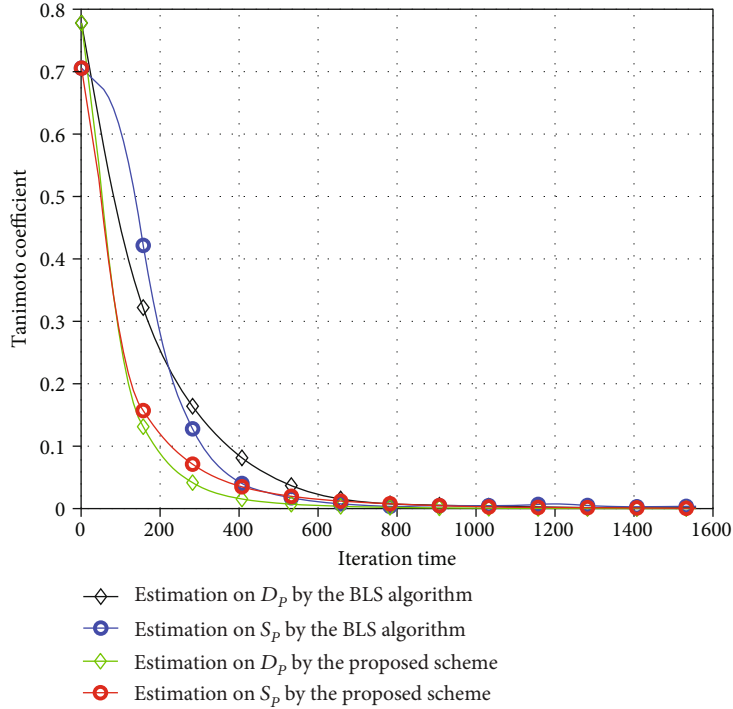
ALGORITHM 2: Grouping Scheme

- ◇ Estimation on D_P by the BLS algorithm
- ○ Estimation on S_P by the BLS algorithm
- ◇ Estimation on D_P by the proposed scheme
- ○ Estimation on S_P by the proposed scheme

FIGURE 2: The Tanimoto coefficient comparison.

TABLE 1: Matching on u_p and v_p.

$<\mathbf{A}(u_p, v_p)\mid\mathbf{Y}>$	$v_1(0.6121)$	$v_2(0.3991)$	$v_3(0.3363)$
$u_1(0.9061)$	18.9036	24.5525	18.9036
$u_2(0.7071)$	63.1738	15.0303	2.6220
$u_3(0.1741)$	6.3134	19.4114	22.7863

TABLE 2: Estimation on u_p, v_p, and $\beta(\theta_p, \phi_p)$.

p	1	2	3
Realθ_p	25.0000°	45.0000°	80.0000°
Estimation of θ_p	25.0282°	45.0005°	79.9737°
Realϕ_p	20.0000°	30.0000°	70.0000°
Estimation of ϕ_p	19.3768°	30.0451°	70.0308°
Real$\beta(\theta_p, \phi_p)$	0.6000	0.8000	1.0000
Estimation of $\beta(\theta_p, \phi_p)$	0.5993	0.8045	1.0011

The principle of the scheme is to guarantee that the UTs in the same group do not have overlapping paths. For example, there are two UTs, u_1 and u_2 with the DoA angles of θ_{u_1}, θ_{u_2} and ϕ_{u_1}, ϕ_{u_2} estimated in the initial stage. If the minimum difference on the DoA angles between u_1 and u_2 is bigger than a threshold denoted as λ, $(\min\{|\theta_{u_1} - \theta_{u_2}|, |\phi_{u_1} - \phi_{u_2}|\} \geq \lambda)$, they would be put into the same group. Based on this rule, the UTs can be designated into different groups, and the UTs in the same group have different distinguishable paths. As a result, the UTs in the same group can use the same training sequence, and their CSI can be estimated and tracked by Algorithm 1.

Algorithm 2: Grouping Scheme1: Initialization: divide UTs into $\lceil U/I \rceil$ groups. Different UTs in the same group are allocated with different pilot sequences;

2: UTs in the same group transmit the pilots in the same time slot. The BS distinguishes the received signals from different UTs;

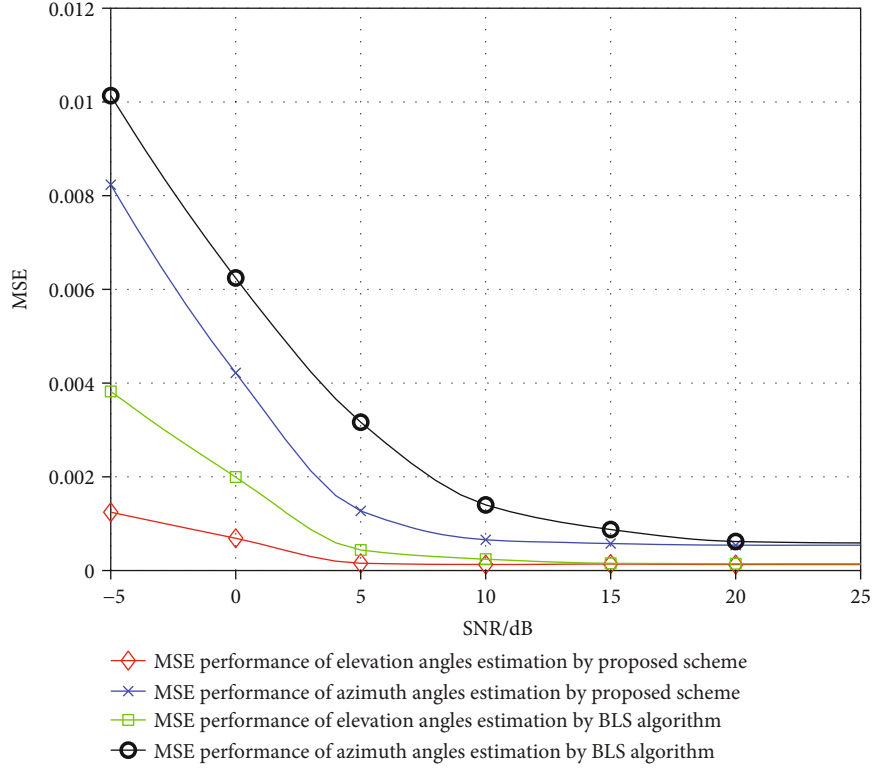3: The BS estimates the initial DoA angles of different UTs;

FIGURE 3: The MSE on DoA estimation.

4: If the minimum difference on the DoA angles between any two UTs is bigger than a threshold denoted as λ, they are put into the same group;

5: The UTs in the same group will use the same pilot sequence.

With the grouping scheme proposed in Algorithm 2, the channel estimation scheme proposed in Algorithm 1 can be extended to the multi-UT scenario. First, the UT grouping scheme is executed at the BS. Then, the UTs in the same group will use the same pilot sequence and transmit simultaneously. After the BS receives signals from the UTs in the same group, it updates the covariance matrix based on (18). Then, it estimates the channel state information based on Algorithm 1 proposed in Section 4. It is noteworthy that the BS treats the UTs in the same group as one "virtual UT." Since there exists a certain gap in DoA angles among UTs in the same group, the paths associated with their channels can be distinguished at the BS. Based on the initial DoA estimation, the BS can identify to which the UT paths belong. When the BS cannot track the CSI accurately, it needs to rerun the grouping scheme in the algorithm. To guarantee the channel estimation and tracking accuracy, the BS may run Algorithm 2 after certain time slots in practice.

6. Numerical Results

In this section, we provide the numerical results to demonstrate the superiority of the proposed scheme. In the simula-

tion setup, the number of antennas deployed at the BS on both elevation and azimuth directions is 64 ($M = N = 64$), and the interantenna spacing, d, is set to be $\lambda/2$. In the first scenario, we assume that there is a single UT under coverage of the BS and the received SNR is 10 dB. Moreover, we assume that the propagation from the UT to BS goes three resolvable paths. The elevation and azimuth angles corresponding to three paths are (25°, 45°, 80°) and (20°, 30°, 70°), respectively.

6.1. The Estimation Accuracy and Convergence Speed. First, to evaluate the estimation accuracy, a Tanimoto coefficient is defined which is given by

$$\varepsilon(k) = 1 - \frac{1}{P} \sum_{i=1}^{P} \frac{\mathbf{t}_{1,i,k} \cdot \mathbf{s}_i}{\left\|\mathbf{t}_{1,i,k}\right\|^2 + \left\|\mathbf{s}_i\right\|^2 - \mathbf{t}_{1,i,k} \cdot \mathbf{s}_i}, \qquad (46)$$

where $\mathbf{s}_i$ is the ith eigenvector of matrix $\mathbf{S}_P$, and $\mathbf{t}_{1,i,k}$ is the ith eigenvector of the estimation matrix $\mathbf{S}_P$ at the kth iteration step. Tanimoto coefficient defined in (46) implies that if ε tends to be 0, the estimation approaches to the real $\mathbf{S}_P$. To demonstrate the superiority of the optimal step size derived in Section 3.4, the Backtracking Line Search (BLS) method proposed in [25] is used for comparison. From Figure 2, we can observe that ε achieved by the PCA-based algorithm and BLS algorithm converges to 0, which indicates the availability of these algorithm on estimation of eigenvector matrix. Meanwhile, during the first 400 iterations, the ε achieved by the PCA-based algorithm is less than the BLS
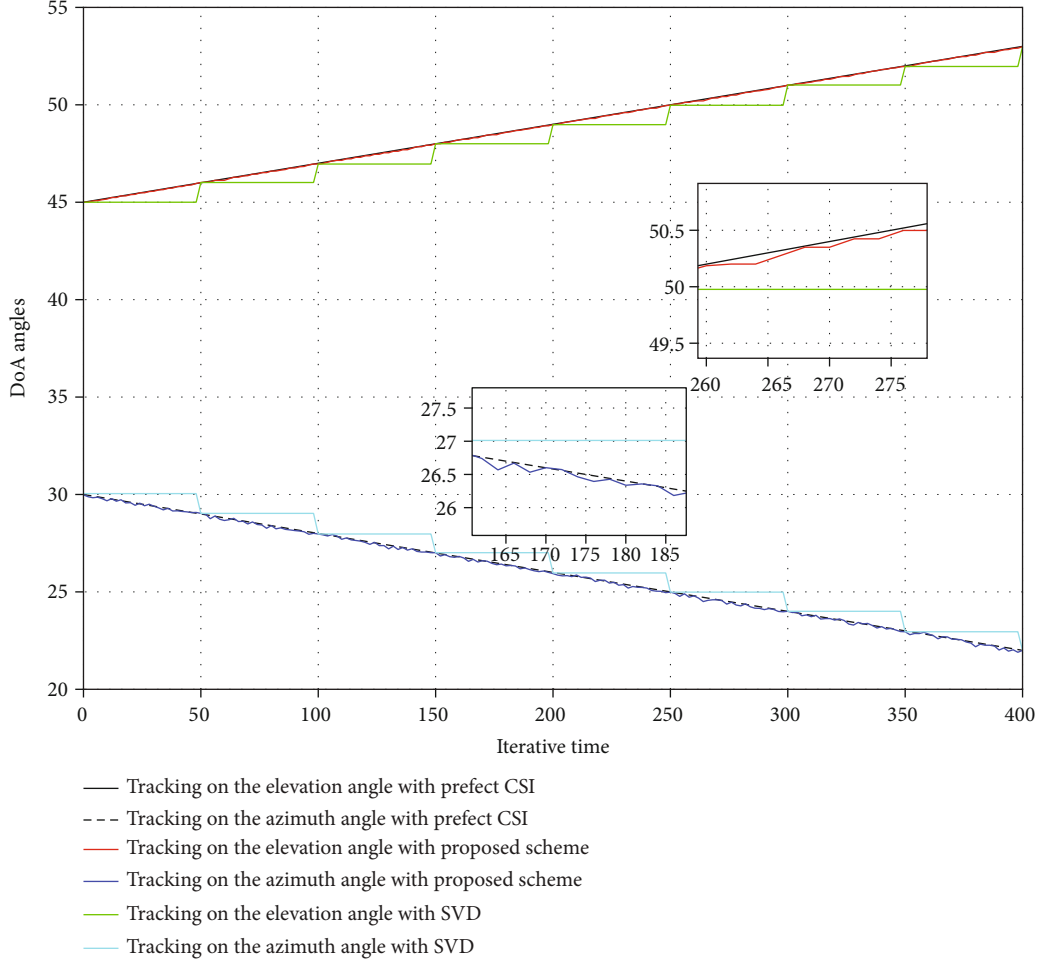
FIGURE 4: Tracking on the DoA angles.

algorithm. The result indicates that the proposed scheme has the high accuracy under the same iterations. From another point of view, the proposed scheme has the fast convergence speed on estimation of $\mathbf{S}_P$.

After recovering the eigenvectors, the intermediate variables, u_p and v_p, are estimated and paired according to the proposed scheme. To demonstrate the match process clearly, the inner products of the received signal matrix on the steering matrix are listed in Table 1, where the steering vectors are constructed by the estimated u_p and v_p.

From Table 1, we can match u_p with the right v_p by finding the maximum value. Consequently, u_1, u_2, and u_3 should be paired with v_2, v_1, and v_3, respectively. After pairing, the estimation of DoA angles and path amplitudes can be obtained through (36), (37), and (43), respectively, which is listed in Table 2, where the index p is to indicate the pth path.

From Table 2, we can observe that the estimations on both θ_p, ϕ_p, and $\beta(\theta_p, \phi_p)$ are close to the real values, which verifies the effectiveness of the proposed scheme. Moreover, since the ϕ_p is determined by both u_p and v_p, the estimation error on u_p and v_p would affect the estimation accuracy on ϕ_p. Therefore, the estimation accuracy of the azimuth angles, ϕ_p, is less than that on the elevation angles, θ_p.

6.2. The Impact of Received SNR on the DoA Estimation. To reveal estimation performance with different SNRs, the normalized mean square error (MSE) on DoA angle estimation is demonstrated in this section. The MSE is defined as

$$\text{MSE}_\theta = \frac{1}{P} \sum_{p=1}^{P} \left(\frac{|\overline{\theta}_p - \theta_p|}{\theta_p} \right), \tag{47}$$

$$\text{MSE}_\phi = \frac{1}{P} \sum_{p=1}^{P} \left(\frac{|\overline{\phi}_p - \phi_p|}{\phi_p} \right). \tag{48}$$

Figure 3 illustrates the MSE performance on DoA angle estimation. It can be observed that the estimation MSE on DoA angles reduces as the received SNR at the BS increases. Figure 3 also compares the proposed scheme with the BLS algorithm with respect to DoA angle estimation MSE. We can observe that the proposed scheme outperforms the BLS algorithm on the estimation accuracy. Since the estimation on azimuth DoA angle depends on both the estimated u_p and v_p, the estimation accuracy on the elevation DoA angle is higher than that on the azimuth DoA angle.

TABLE 3: Add caption.

The number of UT	1	2	3	4	5	6	7	8	9	10
The corresponding elevation and azimuth angles (θ_p, ϕ_p)	$(25°, 20°)$	$(45°, 30°)$	$(80°, 70°)$	$(37°, 82°)$	$(60°, 38°)$	$(75°, 43°)$	$(16°, 8°)$	$(86°, 53°)$	$(17°, 32°)$	$(53°, 28°)$
	$(55°, 60°)$	$(65°, 10°)$	$(14°, 50°)$	$(50°, 25°)$	$(25°, 19.5°)$	$(31°, 48°)$	$(70°, 55°)$	$(69.5°, 60°)$	$(48°, 14°)$	$(27°, 66°)$

(a) Using GS

(b) Without using GS
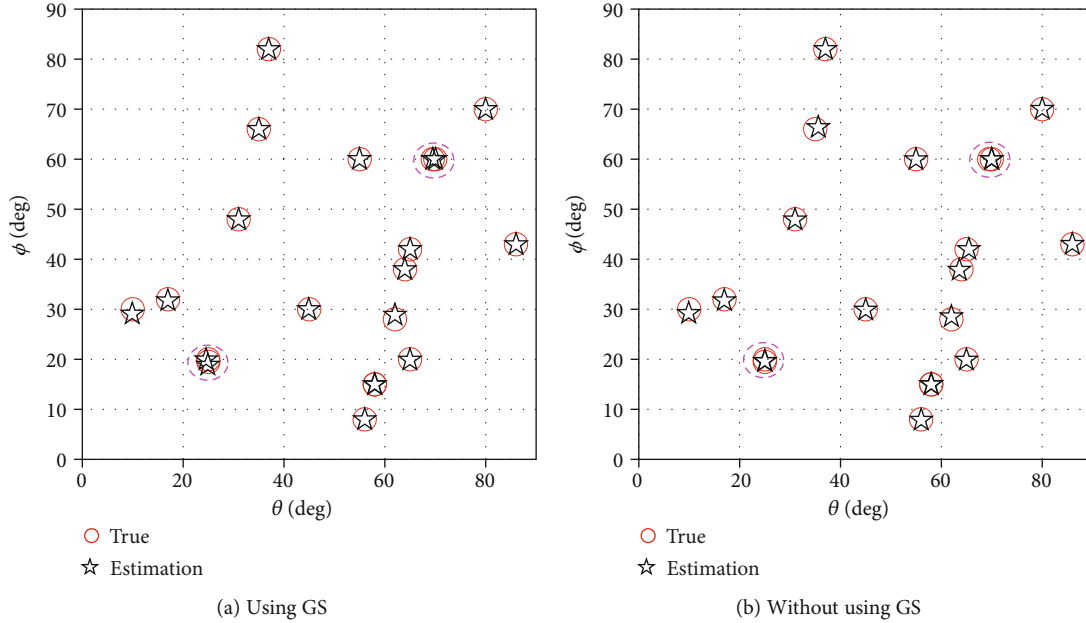
FIGURE 5: (a) When the proposed grouping scheme in Algorithm 2 is adopted, all paths from different UTs can be distinguished at the BS. (b) When the random grouping scheme is adopted, the overlapping paths from different UTs cannot differentiated.

6.3. Tracking Performance. To verify the tracking ability of the proposed scheme, the SVD-based channel estimation is simulated for comparison, where the SVD operation is executed periodically to update the channel estimation. Assuming that the elevation angle changes from 45° to 53°, and azimuth angles change from 30° to 22°, with the changing rate of 0.02°/stepsize.

As shown in Figure 4, the SVD-based channel estimation method cannot track the change on the channel even though it may estimate channel with high accuracy in some times. On the other hand, the channel estimation based on Algorithm 1 is very close to the perfect one all the time even when the channel changes. Therefore, this simulation result confirms the tracking ability on the channel estimation of Algorithm 1.

6.4. The Effectiveness of Grouping Scheme. For the sake of illustration on the effect of the UT grouping scheme, a multiple UT scenario is considered in this section. The elevation and azimuth angles corresponding to the distinguish paths for 10 different UTs are listed in Table 3. Assuming that there are only 2 orthogonal pilot sequences and set $\lambda = 5°$. First, the UTs are assigned to 5 groups. UTs in 5 groups will reuse the 2 orthogonal pilot sequences in a TDMA manner, and the BS can obtain the initial DoA angle estimations of UTs. Afterwards, based on the grouping scheme (GS) in Algorithm 2,

UT1, UT2, UT3, UT4, and UT7 should be put into one group, and the remaining UTs should be put into another group. To better demonstrate the effectiveness of GS, another random grouping scheme is adopted for comparison, where the UT1 to UT5 are put into the same group, and the UT6 to UT10 are put into another group. Accordingly, the estimation on DoA angles is shown in Figure 5.

In Figure 5, the pentagram is used to depict the estimation on DoA angles, the red circle is used to represent the real DoA angles, and the pink dotted circle is used as an indicator. From Figure 5(a), we can observe that the BS can distinguish all paths from different UTs when the proposed GS in Algorithm 2 is adopted. Even the real DoA angles from the two UTs are very close, they can be differentiated by the BS. On the contrary, in Figure 5(b), the pink circle with the dotted line shows that two overlapped circles only contain one corresponding pentagram, which means the system is failed to distinguish the paths between UT1 and UT5. Therefore, when multiple UTs are present in the system, the GS is necessary for the BS to achieve accurate CSI via the proposed channel estimation and tracking scheme.

6.5. Multiple UTs Tracking. Due to the mobility of UTs, tracking the CSI accurately needs to run the GS at the BS periodically. Therefore, the periodicity to execute the GS at
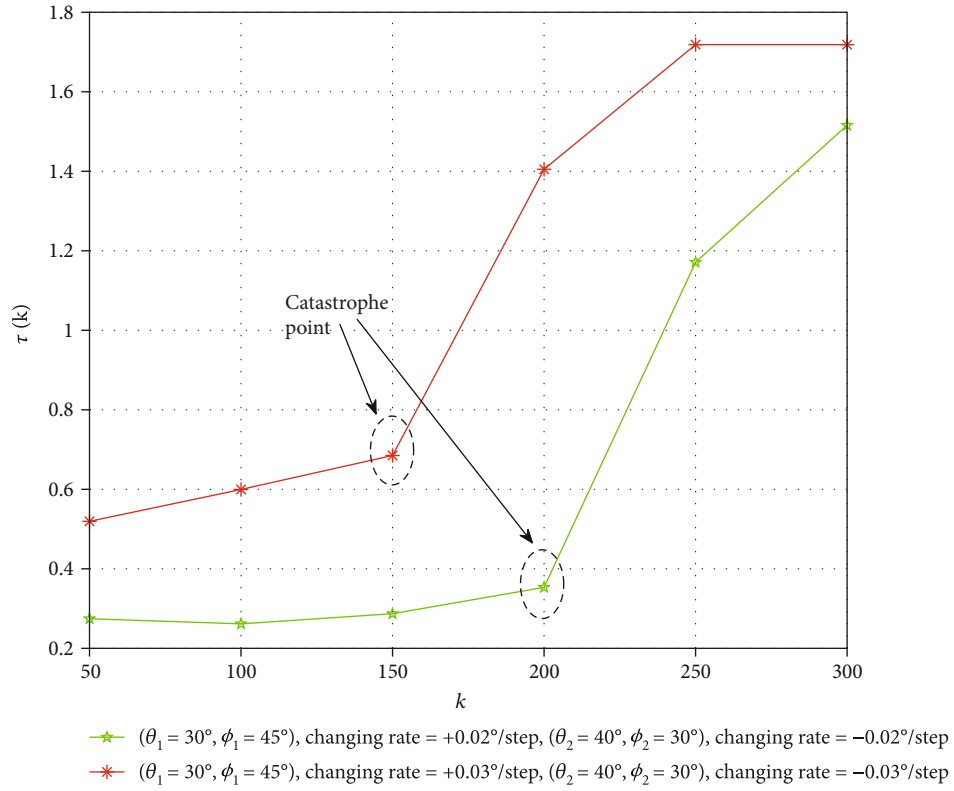
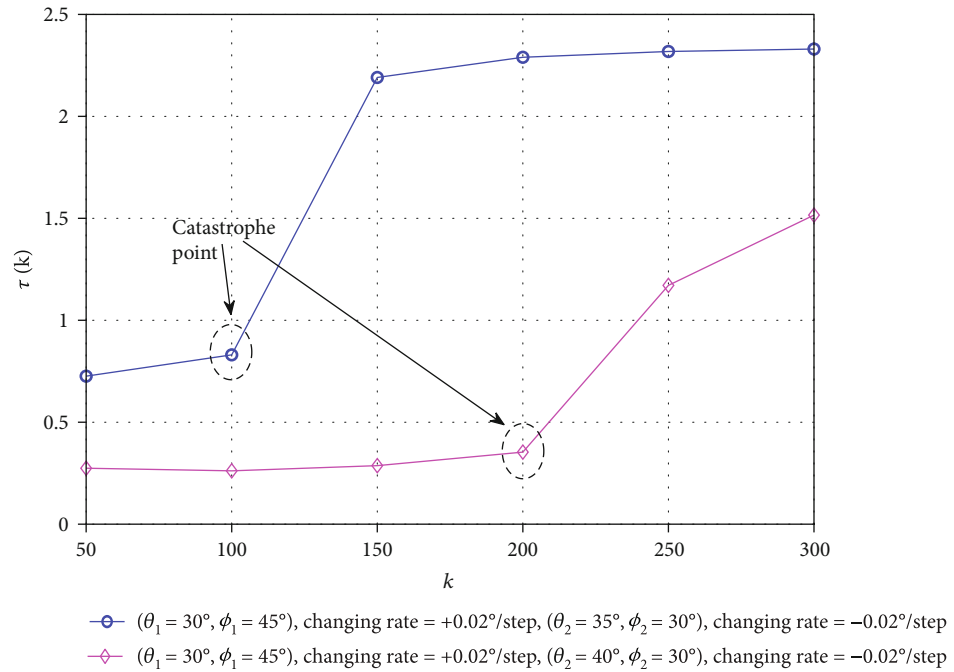Figure 6: The impact of changing rate on the tracking accuracy.

Figure 7: The impact of initial DoA angle on the tracking accuracy.

the BS is the key to guarantee tracking accuracy. Without loss of generality, the scenario with two UTs is simulated in this section. In the simulation setup, we assume that the signal transmission from each UT to the BS will go through the same path. Moreover, different initial DoA angles and step size are simulated for comparison.

Accordingly, the period to execute Algorithm 2 is denoted as k time slots. To evaluate the effect of the period on the tracking accuracy, a coefficient, denoted as τ, is defined as

$$\tau(k) = \sum_{i=1}^{U} \sum_{j=1}^{J} \frac{\left| \widehat{\theta}(k)_{i,j} - \theta(k)_{i,j} \right| + \left| \widehat{\phi}(k)_{i,j} - \phi(k)_{i,j} \right|}{J}, \quad (49)$$

where $\widehat{\theta}(k)_{i,j}$ and $\widehat{\phi}(k)_{i,j}$ are the estimations of elevation and azimuth angles of the ith UT in the jth time slot, respectively; J is the total running time slots, which is set to $J = 600$. $\tau(k)$ implies the average estimation error on the DoA angles on both elevation and azimuth directions.

The tracking accuracy with different periods is illustrated in Figures 6 and 7, which indicates that optimal GS period is directly related to the DoA angle changing rate and the initial DoA angle difference on elevation direction associated to different UTs. We can observe that the tracking accuracy decreases when the period, k, increases. Optimal period for GS is inversely proportional to the changing rate and proportional to the initial DoA angle difference on elevation direction. Moreover, it is noteworthy that the curve with the diamond marker is stationary even if the difference in the azimuth DoA angles between two UTs is small. Actually, the difference in the azimuth angles has a little effect on the tracking accuracy when the elevation angles of UTs are distinguishable.

7. Conclusion

In this paper, we have proposed a CSI estimation and tracking method for millimeter wave massive MIMO system. We have illustrated the CSI estimation from two aspects: computational complexity and pilot contamination. First, to reduce the computational complexity, a PCA-based algorithm is applied to recover the eigenvectors of the received signal at the BS. Then, with estimated eigenvectors, DoA angles can be estimated by using the MUSIC algorithm and inner products operation. To reduce the pilot overhead, we have utilized the grouping method to divide UTs into different groups. Then, UTs in the same group can share the same pilot sequence and transmit simultaneously without causing interference. In order to express the convergence speed of the proposed algorithm, a BLS algorithm is used for comparison. Compared to this algorithm, the proposed algorithm has faster convergence. Meanwhile, most existing methods are one-time estimation schemes, which need to be rerun after the channel coherence time, but the proposed method possesses good tracking behavior. The numerical simulations are established to verify the superior performance of the proposed method.

Notations

$(\mathbf{X})^{\mathrm{T}}$:   The transpose of a matrix X
$(\mathbf{X})^{*}$:   The conjugate of a matrix X
$(\mathbf{X})^{\mathrm{H}}$:   The Hermitian of a matrix X
$\mathbb{E}\{\cdot\}$:   The statistical expectation
$\det(\mathbf{X})$:   The determinant of matrix X
$tr(\mathbf{X})$:   The trace of matrix X
$\|\cdot\|$:   The Euclidian norm
$\lceil\cdot\rceil$:   The maximum integer operator
$\mathbf{I}$:   The identity matrix
$\mathbb{C}$:   The complex domain
ε: The Tanimoto coefficient
θ, ϕ: The DoA angle
$\mathbf{Y}$:   The received signals
$\mathbf{H}$:   The channel matrix
$\mathbf{C}$:   The covariance matrix
$<\mathbf{J}, \mathbf{K}>$:   The standard inner product of two matrix $\mathbf{J}$ and $\mathbf{K}$.

Data Availability

No data available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[2] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[4] E. Bjornson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[5] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[6] S. S. Ioushua and Y. C. Eldar, "A family of hybrid analog-digital beamforming methods for massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 12, pp. 3243–3257, 2019.

[7] T. L. Marzetta, G. Caire, M. Debbah, I. Chih-Lin, and S. K. Mohammed, "Special issue on massive MIMO," *Journal of*

Communications and Networks, vol. 15, no. 4, pp. 333–337, 2013.

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

[9] Y. Chen, L. You, X. Gao, and X. G. Xia, "Channel estimation with pilot reuse in IQ imbalanced massive MIMO," *IEEE Access*, vol. 8, pp. 1542–1555, 2020.

[10] L. Ku, J. Fan, and J. Deng, "Low complexity pilot allocation in massive MIMO systems," in *2016 8th IEEE International Conference on Communication Software and Networks*, pp. 402–406, 2016.

[11] J.-C. Chen, C.-K. Wen, S. Jin, and K.-K. Wong, "A low complexity pilot scheduling algorithm for massive MIMO," *IEEE Wireless Communications Letters*, vol. 6, no. 1, pp. 18–21, 2017.

[12] J. Fan, W. Li, and Y. Zhang, "Pilot contamination mitigation by fractional pilot reuse with threshold optimization in massive MIMO systems," *Digital Signal Processing*, vol. 78, pp. 197–204, 2018.

[13] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3170–3184, 2017.

[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 Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1948–1961, 2017.

[15] A. Wang, R. Yin, and C. Zhong, "Channel estimation for uniform rectangular array based massive MIMO systems with low complexity," *IEEE Transactions on Vehicular Technology*, vol. 99, no. 99, pp. 1–12, 2019.

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

[17] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, 2016.

[18] W. Zhang, T. Kim, and S. H. Leung, "A sequential subspace method for millimeter wave MIMO channel estimation," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5355–5368, 2020.

[19] S. Haghighatshoar and G. Caire, "Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8316–8332, 2017.

[20] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, "Fast channel tracking for terahertz beamspace massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 5689–5696, 2017.

[21] Y. Yang, F. Gao, G. Y. Li, and M. Jian, "Deep learning-based downlink channel prediction for FDD massive MIMO system," *IEEE Communications Letters*, vol. 23, no. 11, pp. 1994–1998, 2019.

[22] R. Yin, X. Zhou, A. Wang, C. Zhong, C. Wu, and X. Chen, "Adaptive channel estimation and tracking for URA-based massive MIMO systems," *IEEE Access*, vol. 8, pp. 54213–54224, 2020.

[23] L. Xu, "Least mean square error reconstruction principle for self-organizing neural- nets," *Neural Networks*, vol. 6, no. 5, pp. 627–648, 1993.

[24] J. P. Delmas, *Adaptive Signal Processing: Next Generation Solutions-Subspace Tracking for Signal Processing*, Wiley-IEEE Press, 2010.

[25] T. Mikosch, S. Resnick, and S. Robinson, *Numerical Optimization*, Springer Science+Business Media, LLC, Publication, 2002.