Over-sampled Beamspace Channel Estimation for Millimeter Wave Massive MIMO

Wenyan Ma and Chenhao Qi
School of Information Science and Engineering
Southeast University, Nanjing 210096, China
Email: qch@seu.edu.cn

***Abstract*—Beamspace channel estimation for millimeter wave (mmWave) massive MIMO system is investigated. An identity matrix approximation (IA)-based beamspace channel estimation scheme is proposed, including the design of hybrid precoding and combining matrix as well as searching the largest entry of over-sampled beamspace receiving matrix. The design of hybrid combining and hybrid precoding is formulated as two optimization problems. By decoupling the design of analog combining and digital combining, closed-form solutions are obtained. Then an algorithm based on bisection search is proposed to search the largest entry of the over-sampled beamspace receiving matrix. Additionally, computational complexity is compared between the proposed scheme and the existing channel estimation schemes. Simulation results show that the proposed IA-based beamspace channel estimation scheme outperforms the existing schemes.**

***Index Terms*—Millimeter wave communications, channel estimation, hybrid precoding, massive MIMO, beamspace.**

I. Introduction

Millimeter wave (mmWave) communications, which range from 30GHz to 300GHz in frequency, have recently attracted great attention due to high spectrum efficiency and large available bandwidth [1]. Due to the large number of antennas at both transmitting and receiving sides, the size of channel matrix is very large, therefore the channel estimation is rather time consuming. To reduce the complexity of channel estimation in mmWave massive MIMO system, some advanced schemes based on the beamspace channel have been proposed very recently [2], [3]. The key idea of these schemes is to efficiently explore the sparsity of beamspace channel by sparse signal processing techniques.

However, the sparsity of beamspace channel may be impaired by power leakage due to the limited beamspace resolution, indicating that the beamspace channel is not ideally sparse and there are many small nonzero entries [4]. Therefore, it brings extra challenge for the sparse recovery [5]. To solve this problem, a discrete compressive sensing (DCS)-based channel estimation scheme is proposed in [4], where the beamspace resolution can be set arbitrarily and can be higher than the reciprocal of the antenna number, leading to improved channel estimation performance. Moreover, an over-sampled compressive sensing (OCS)-based channel estimation scheme is proposed in [5], where the measurement matrix is consisted of a large number of over-sampled steering vectors and is capable of estimating the angle of arrival (AoA) and angle of departure (AoD) more accurate than conventional measurement matrix. However, both DCS-based and OCS-based schemes use random measurement matrices, which can not always achieve the optimal performance. In particular, the estimated AoA and AoD can not be always within the scale of quantization error even without any noise. Therefore deterministic measurement matrix is more appealing.

In this paper, we investigate the beamspace channel estimation for mmWave massive MIMO system. We propose an identity matrix approximation (IA)-based beamspace channel estimation scheme, which includes the design of the hybrid precoding and combining matrix as well as searching the largest entry of over-sampled beamspace receiving matrix. We formulate the design of hybrid combining and hybrid precoding as two optimization problems and then obtain closed-form solutions by decoupling the design of analog combining and digital combining. After that, we propose an algorithm to search the largest entry of the over-sampled beamspace receiving matrix. The algorithm is based on bisection search and includes two stages, where we find the main lobe in the first stage, and we find the largest entry corresponding to the peak within the main lobe in the second stage. Additionally, we compare the computational complexity between the proposed scheme and the existing channel estimation schemes.

The notations are defined as follows. Symbols for matrices (upper case) and vectors (lower case) are in boldface. $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^{-1}$, $\boldsymbol{I}_L$, $\mathbf{1}_L$, $\mathbb{C}^{M\times N}$, $\otimes$, $\text{vec}(\cdot)$, $\text{E}\{\cdot\}$, $\mathbf{0}^M$, $\|\cdot\|_0$, $\|\cdot\|_2$, $\|\cdot\|_F$, $\boldsymbol{A}[p,q]$, $\langle\cdot\rangle$, $\text{Tr}(\cdot)$ and $\mathcal{CN}$, denote the transpose, conjugate transpose (Hermitian), inverse, identity matrix of size L, vector of size L with all entries being 1, the set of $M \times N$ complex-valued matrices, kronecker product, vectorization, expectation, zero vector of size M, l_0-norm, l_2-norm, Frobenius norm, entry of $\boldsymbol{A}$ at the pth row and qth column, round, trace and complex Gaussian distribution, respectively.

II. Problem Formulation

We consider an uplink multiuser mmWave massive MIMO system comprising a BS and U users. Both the BS and users are equipped with an uniform linear array (ULA) [6]. Denote N_A, M_A, N_R and M_R as the number of antennas at the BS, number of antennas at each user, number of RF chains at the BS and number of RF chains at each user. In practical mmWave massive MIMO with hybrid precoding and combining, the number of RF chains is much smaller than that of antennas, i.e., $N_R \ll N_A$ and $M_R \ll M_A$. Each user performs analog precoding in RF and digital precoding in the baseband, while the BS performs analog combining in RF and digital combining in the baseband for uplink transmission. The received signal vector at the BS can be represented as

$$\boldsymbol{y} = \boldsymbol{W}_B \boldsymbol{W}_R \sum_{u=1}^{U} \boldsymbol{H}_u \boldsymbol{F}_{R,u} \boldsymbol{F}_{B,u} \boldsymbol{s}_u + \boldsymbol{W}_B \boldsymbol{W}_R \boldsymbol{n} \quad (1)$$

where $\boldsymbol{F}_{B,u} \in \mathbb{C}^{M_R \times M_R}$, $\boldsymbol{F}_{R,u} \in \mathbb{C}^{M_A \times M_R}$, $\boldsymbol{W}_B \in \mathbb{C}^{N_R \times N_R}$, and $\boldsymbol{W}_R \in \mathbb{C}^{N_R \times N_A}$ are the digital precoding matrix, analog precoding matrix, digital combining matrix, and analog combining matrix for the $u(u = 1, 2, \ldots, U)$th user, respectively. $\boldsymbol{s}_u \in \mathbb{C}^{M_R}$ denotes the signal vector satisfying $\mathrm{E}\{\boldsymbol{s}_u \boldsymbol{s}_u^H\} = \boldsymbol{I}_{M_R}$. $\boldsymbol{n} \in \mathbb{C}^{N_A}$ denotes additive white Gaussian noise (AWGN) vector satisfying $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma^2 \boldsymbol{I}_{N_A})$. $\boldsymbol{H}_u \in \mathbb{C}^{N_A \times M_A}$ denotes the channel matrix between the BS and the uth user and can be expressed according to the widely used Saleh-Valenzuela channel model [1] as

$$\boldsymbol{H}_u = \sqrt{\frac{N_A M_A}{L_u}} \sum_{i=1}^{L_u} g_{u,i} \boldsymbol{\alpha}(N_A, \theta_{u,i}) \boldsymbol{\alpha}^H(M_A, \varphi_{u,i}) \quad (2)$$

where L_u and $g_{u,i}$ denote the total number of resolvable paths and the channel fading coefficient of the ith path, respectively. Usually, there is a strong line-of-sight (LOS) path ($i = 1$) and $L_u - 1$ much weaker non-line-of-sight (NLOS) paths ($2 \leq i \leq L_u$). The mmWave transmission essentially relies on the strong LOS path. The steering vector $\boldsymbol{\alpha}(N, \theta)$ is defined as $\boldsymbol{\alpha}(N, \theta) = \left[1, e^{-j\pi\theta}, ..., e^{-j\pi\theta(N-1)}\right]^T / N$. Define the AoA and AoD of the ith path of the uth user as $\Theta_{u,i}$ and $\Phi_{u,i}$, respectively. Further define $\theta_{u,i} \triangleq \frac{2d_{BS}}{\lambda} \sin \Theta_{u,i}$ and $\varphi_{u,i} \triangleq \frac{2d_{UE}}{\lambda} \sin \Phi_{u,i}$, where d_{BS} and d_{UE} denote the antenna interval of the BS and users, respectively. We usually set $d_{BS} = d_{UE} = \lambda/2$, where λ is the wavelength of mmWave signal. In practice, both $\theta_{u,i}$ and $\varphi_{u,i}$ obey the uniform distribution [-1,1] [7].

It is observed that $\boldsymbol{y}$ in (1) is a mixed signal from different users. To distinguish different user signal at the BS, each user repeatedly transmits an orthogonal pilot sequence $\boldsymbol{p}_u \in \mathbb{C}^U$ with length of U for $T_1 T_2$ times. For simplicity, we suppose each user transmit the same pilot sequence for all M_R RF chains, where the pilot matrix for the uth user can be defined as $\boldsymbol{P}_u \triangleq [\boldsymbol{p}_u, \boldsymbol{p}_u, \ldots, \boldsymbol{p}_u]^H = \mathbf{1}_{M_R} \boldsymbol{p}_u^H \in \mathbb{C}^{M_R \times U}$. The channel keeps constant during $T \triangleq T_1 T_2 U$ time slots [8]. We use T_1 different digital precoding matrix and analog precoding matrix, denoted as $\boldsymbol{F}_{B,u}^{t_1} \in \mathbb{C}^{M_R \times M_R}$ and $\boldsymbol{F}_{R,u}^{t_1} \in \mathbb{C}^{M_A \times M_R}$, respectively, $t_1 = 1, 2, \ldots, T_1$, at the $u(u = 1, 2, \ldots, U)$th user. We use T_2 different digital combining matrix and analog combining matrix, denoted as $\boldsymbol{W}_B^{t_2} \in \mathbb{C}^{N_R \times N_R}$ and $\boldsymbol{W}_R^{t_2} \in \mathbb{C}^{N_R \times N_A}$, respectively, $t_2 = 1, 2, \ldots, T_2$, at the BS. During the T_1 repetitive transmission of pilot sequence from the $((t_2 - 1)T_1 + 1)$th transmission to $(t_2 T_1)$th transmission, we use T_1 different $\boldsymbol{F}_{B,u}^{t_1}$ and $\boldsymbol{F}_{R,u}^{t_1}$ for hybrid precoding while use the same $\boldsymbol{W}_B^{t_2}$ and $\boldsymbol{W}_R^{t_2}$ for hybrid combining, where the received pilot matrix $\boldsymbol{Y}^{t_1,t_2} \in \mathbb{C}^{N_R \times U}$ can be denoted as

$$\boldsymbol{Y}^{t_1,t_2} = \boldsymbol{W}_B^{t_2} \boldsymbol{W}_R^{t_2} \sum_{u=1}^{U} \boldsymbol{H}_u \boldsymbol{F}_{R,u}^{t_1} \boldsymbol{F}_{B,u}^{t_1} \boldsymbol{P}_u + \boldsymbol{W}_B^{t_2} \boldsymbol{W}_R^{t_2} \boldsymbol{N}^{t_1,t_2}$$

with $\boldsymbol{N}^{t_1,t_2} \in \mathbb{C}^{N_A \times U}$ representing the AWGN matrix. Each entry of $\boldsymbol{N}^{t_1,t_2}$ independently obeys complex Gaussian distribution with zero mean and variance of σ^2. To ease the notation, we define $\widetilde{\boldsymbol{N}}^{t_1,t_2} \triangleq \boldsymbol{W}_B^{t_1} \boldsymbol{W}_R^{t_1} \boldsymbol{N}^{t_1,t_2}$. Due to the orthogonality of $\boldsymbol{p}_u$, i.e., $\boldsymbol{p}_u^H \boldsymbol{p}_u$=1 and $\boldsymbol{p}_u^H \boldsymbol{p}_i = 0$, $\forall u, i \in \{1, 2, \ldots, U\}$, $i \neq u$ [3], we can obtain the measurement vector $\boldsymbol{r}_u^{t_1,t_2} \in \mathbb{C}^{N_R}$ for the uth user by multiplying $\boldsymbol{Y}^{t_1,t_2}$ with $\boldsymbol{p}_u$ as

$$\boldsymbol{r}_u^{t_1,t_2} = \boldsymbol{Y}^{t_1,t_2} \boldsymbol{p}_u = \boldsymbol{W}^{t_2} \boldsymbol{H}_u \boldsymbol{f}_u^{t_1} + \widetilde{\boldsymbol{n}}^{t_1,t_2} \quad (3)$$

where

$$\boldsymbol{W}^{t_2} \triangleq \boldsymbol{W}_B^{t_2} \boldsymbol{W}_R^{t_2}, \boldsymbol{f}_u^{t_1} \triangleq \boldsymbol{F}_{R,u}^{t_1} \boldsymbol{F}_{B,u}^{t_1} \mathbf{1}_{M_R}, \widetilde{\boldsymbol{n}}^{t_1,t_2} \triangleq \widetilde{\boldsymbol{N}}^{t_1,t_2} \boldsymbol{p}_u. \quad (4)$$

Define $T_3 \triangleq T_2 N_R$. We stack the T_2 received pilot sequences together and have $\boldsymbol{r}_u^{t_1} = \boldsymbol{W} \boldsymbol{H}_u \boldsymbol{f}_u^{t_1} + \widetilde{\boldsymbol{n}}^{t_1}$, where

$$\begin{aligned} \boldsymbol{r}_u^{t_1} &\triangleq [(\boldsymbol{r}_u^{t_1,1})^T, (\boldsymbol{r}_u^{t_1,2})^T, \ldots, (\boldsymbol{r}_u^{t_1,T_2})^T]^T \in \mathbb{C}^{T_3}, \\ \boldsymbol{W} &\triangleq [(\boldsymbol{W}^1)^T, (\boldsymbol{W}^2)^T, \ldots, (\boldsymbol{W}^{T_2})^T]^T \in \mathbb{C}^{T_3 \times N_A}, \\ \widetilde{\boldsymbol{n}}^{t_1} &\triangleq [(\widetilde{\boldsymbol{n}}^{t_1,1})^T, (\widetilde{\boldsymbol{n}}^{t_1,2})^T, \ldots, (\widetilde{\boldsymbol{n}}^{t_1,T_2})^T]^T \in \mathbb{C}^{T_3}. \end{aligned} \quad (5)$$

Further define

$$\boldsymbol{R}_u \triangleq [\boldsymbol{r}_u^1, \boldsymbol{r}_u^2, \ldots, \boldsymbol{r}_u^{T_1}], \ \boldsymbol{F}_u \triangleq [\boldsymbol{f}_u^1, \boldsymbol{f}_u^2, \ldots, \boldsymbol{f}_u^{T_1}], \\ \widetilde{\boldsymbol{n}} \triangleq [\widetilde{\boldsymbol{n}}^1, \widetilde{\boldsymbol{n}}^2, \ldots, \widetilde{\boldsymbol{n}}^{T_1}]. \quad (6)$$

We have

$$\boldsymbol{R}_u = \boldsymbol{W} \boldsymbol{H}_u \boldsymbol{F}_u + \widetilde{\boldsymbol{n}}. \quad (7)$$

Considering the LOS channel path concentrates most channel power in mmWave massive MIMO systems, we usually use LOS channel path to transmit data for the uth user [4]. Therefore, it is important to design $\boldsymbol{W}$ and $\boldsymbol{F}_u$ to estimate the AoA and the AoD of the LOS path of $\boldsymbol{H}_u$ in (7) for the uth user, which will be discussed in the following sections.

III. Over-sampled Beamspace Channel Estimation

We first propose a framework of over-sampled beamspace channel estimation for mmWave massive MIMO system. Then based on this framework, we propose an IA-based channel estimation scheme. Finally, we compare the computational complexity.

A. Framework of Over-sampled Beamspace Channel Estimation

Denote $\bar{\boldsymbol{H}}_u^v \in \mathbb{C}^{N_A \times M_A}$ as the beamspace channel matrix for the uth user, which can be represented as [3]

$$\bar{\boldsymbol{H}}_u^v = \boldsymbol{D}(N_A, N_A)^H \boldsymbol{H}_u \boldsymbol{D}(M_A, M_A) \quad (8)$$

where $\boldsymbol{D}(N, K) \in \mathbb{C}^{N \times K}$ is the sampling matrix, which is defined as $\boldsymbol{D}(N, K) \triangleq [\boldsymbol{\alpha}(N, -1 + 0/K), \boldsymbol{\alpha}(N, -1 + 2/K), \boldsymbol{\alpha}(N, -1 + 4/K), \ldots, \boldsymbol{\alpha}(N, -1 + 2(K-1)/K)]$. In fact, $\boldsymbol{D}(N, K)$ samples the beamspace [-1,1] in an interval of $2/K$ by K steering vectors. For $\bar{\boldsymbol{H}}_u^v$, the AoA and AoD is sampled in an interval of $2/N_A$ and $2/M_A$, respectively. Therefore the quantization error for the estimated AoA and AoD is $2/N_A$ and $2/M_A$, respectively. In order to decrease the quantization error, we introduce the over-sampled beamspace channel matrix for the uth user $\boldsymbol{H}_u^v \in \mathbb{C}^{K\times K}$ as

$$\boldsymbol{H}_u^v = \boldsymbol{D}(N_A, K)^H \boldsymbol{H}_u \boldsymbol{D}(M_A, K) \tag{9}$$

where K is the number of steering vectors with $K > N_A$ and $K > M_A$. Note that the coordinates of the largest entry of $\boldsymbol{H}_u^v$ are the AoA and AoD of the LOS path with the quantization error of $2/K$. To reduce the quantization error and improve channel estimation, we can use a large K by finding the largest entry of $\boldsymbol{H}_u^v$.

However, we cannot directly obtain $\boldsymbol{H}_u^v$ based on $\boldsymbol{R}_u$ in (7), due to the hybrid precoding and combining operations of $\boldsymbol{F}_u$ and $\boldsymbol{W}$, respectively. Note that the dimension of $\boldsymbol{H}_u$ is $N_A \times M_A$, while the dimension of $\boldsymbol{W}\boldsymbol{H}_u\boldsymbol{F}_u$ is $T_3 \times T_1$. In order to obtain the over-sampled beamspace channel matrix as described in (9), we multiply $\boldsymbol{R}_u$ with $\boldsymbol{W}^H$ on the left and $\boldsymbol{F}_u^H$ on the right, which can make the dimension of $\boldsymbol{W}^H\boldsymbol{W}\boldsymbol{H}_u\boldsymbol{F}_u\boldsymbol{F}_u^H$ the same as that of $\boldsymbol{H}_u$. Now we can obtain an over-sampled beamspace receiving matrix $\boldsymbol{R}_u^v \in \mathbb{C}^{K\times K}$ as

$$\boldsymbol{R}_u^v = \boldsymbol{D}(N_A, K)^H \boldsymbol{W}^H \boldsymbol{W} \boldsymbol{H}_u \boldsymbol{F}_u \boldsymbol{F}_u^H \boldsymbol{D}(M_A, K) + \widetilde{\boldsymbol{n}}^v \tag{10}$$

where $\widetilde{\boldsymbol{n}}^v \triangleq \boldsymbol{D}(N_A, K)^H \boldsymbol{W}^H \widetilde{\boldsymbol{n}} \boldsymbol{F}_u^H \boldsymbol{D}(M_A, K)$. It is expected that

$$\begin{aligned} \boldsymbol{D}(N_A, K)^H \boldsymbol{W}^H \boldsymbol{W} &= \gamma_N \boldsymbol{D}(N_A, K)^H, \\ \boldsymbol{F}_u \boldsymbol{F}_u^H \boldsymbol{D}(M_A, K) &= \gamma_M \boldsymbol{D}(M_A, K). \end{aligned} \tag{11}$$

where $\gamma_N \triangleq \|\boldsymbol{W}^H\boldsymbol{W}\|_F/\sqrt{N_A}$ and $\gamma_M \triangleq \|\boldsymbol{F}_u\boldsymbol{F}_u^H\|_F/\sqrt{M_A}$. However, in this case, it requires that $T_3 \geq N_A$ and $T_1 \geq M_A$, leading to huge pilot overhead. Therefore, we need to reduce the pilot overhead in practice, where none of $\boldsymbol{D}(N_A, K)^H\boldsymbol{W}^H\boldsymbol{W} = \gamma_N \boldsymbol{D}(N_A, K)^H$ and $\boldsymbol{F}_u\boldsymbol{F}_u^H\boldsymbol{D}(M_A, K) = \gamma_M \boldsymbol{D}(M_A, K)$ can be satisfied. Now we have the following two subproblems.

Subproblem 1 *(Hybrid Precoding and Combining Matrix Design):* $\boldsymbol{W}$ and $\boldsymbol{F}_u$ should be well designed so that the coordinates of the largest entry of $\boldsymbol{R}_u^v$ are the AoA and AoD of the LOS path with the quantization error of $2/K$. Therefore, the AoA and AoD of the LOS path can be estimated by finding the largest entry of $\boldsymbol{R}_u^v$.

Subproblem 2 *(Search the Largest Entry):* Due to the large dimension of $\boldsymbol{R}_u^v$, finding the largest entry of $\boldsymbol{R}_u^v$ is computationally expensive. Therefore, it is better to design a low-complexity search algorithm fully regarding the structure of $\boldsymbol{R}_u^v$.

B. Hybrid Precoding and Combining Matrix Design

It is seen that $\boldsymbol{R}_u^v$ in (10) is the over-sampled beamspace channel matrix with noise if (11) is satisfied. However, due to the fact that $T_3 < N_A$, $T_1 < M_A$, the rank of $\boldsymbol{W}^H\boldsymbol{W}$ and $\boldsymbol{F}_u\boldsymbol{F}_u^H$ is less than N_A and M_A, respectively. So we cannot find a proper $\boldsymbol{W}$ and $\boldsymbol{F}_u$ satisfying (11).

The optimization problem for **hybrid combining matrix design** is

$$\begin{aligned} \min_{\boldsymbol{W}} \quad & \|\boldsymbol{D}(N_A, K)^H \boldsymbol{W}^H \boldsymbol{W} - \gamma_N \boldsymbol{D}(N_A, K)^H\|_F \\ \textbf{s.t.} \quad & \boldsymbol{W}_R^{t_2} \in \mathcal{W}_{\mathcal{R}}, \|\boldsymbol{W}_B^{t_2}\boldsymbol{W}_R^{t_2}\|_F^2 = P_W,\ t_2 = 1, 2, \ldots, T_2, \end{aligned} \tag{12}$$

where $\mathcal{W}_{\mathcal{R}}$ is the set of all feasible analog combining matrix and P_W is the given power for the hybrid combining. Note that $\boldsymbol{D}(N_A, K)\boldsymbol{D}(N_A, K)^H = K\boldsymbol{I}_{N_A}/N_A^2$. We have

$$\begin{aligned} & \|\boldsymbol{D}(N_A, K)^H \boldsymbol{W}^H \boldsymbol{W} - \gamma_N \boldsymbol{D}(N_A, K)^H\|_F^2 \\ &= \|\boldsymbol{D}(N_A, K)^H (\boldsymbol{W}^H\boldsymbol{W} - \gamma_N \boldsymbol{I}_{N_A})\|_F^2 \\ &= K\mathrm{Tr}\Big((\boldsymbol{W}^H\boldsymbol{W} - \gamma_N \boldsymbol{I}_{N_A})^H (\boldsymbol{W}^H\boldsymbol{W} - \gamma_N \boldsymbol{I}_{N_A})\Big)/N_A^2 \\ &= K\|\boldsymbol{W}^H\boldsymbol{W} - \gamma_N \boldsymbol{I}_{N_A}\|_F^2/N_A^2. \end{aligned} \tag{13}$$

Then (12) can be further rewritten as

$$\begin{aligned} \min_{\boldsymbol{W}} \quad & \|\boldsymbol{W}^H\boldsymbol{W} - \gamma_N \boldsymbol{I}_{N_A}\|_F \\ \textbf{s.t.} \quad & \boldsymbol{W}_R^{t_2} \in \mathcal{W}_{\mathcal{R}}, \|\boldsymbol{W}_B^{t_2}\boldsymbol{W}_R^{t_2}\|_F^2 = P_W,\ t_2 = 1, 2, \ldots, T_2. \end{aligned} \tag{14}$$

It is seen that $\boldsymbol{W}$ defined in (5) is a flat matrix where the columns are more than the rows. Therefore, it is infeasible that $\boldsymbol{W}^H\boldsymbol{W}$ equals $\gamma_N \boldsymbol{I}_{N_A}$. Instead, it is important to design $\boldsymbol{W}$ so that $\boldsymbol{W}^H\boldsymbol{W}/\gamma_N$ approximates the identity matrix, a.k.a, identity matrix approximation (IA). To minimize $\|\boldsymbol{W}^H\boldsymbol{W} - \gamma_N \boldsymbol{I}_{N_A}\|_F$, $\boldsymbol{W}$ can be a submatrix of $\sqrt{\gamma_N}\boldsymbol{U}$ by selecting the first T_3 rows of $\sqrt{\gamma_N}\boldsymbol{U}$, where $\boldsymbol{U}$ is any $N_A \times N_A$ unitary matrix [9]. For example, we obtain $\boldsymbol{U}$ by singular value decomposition (SVD) of a $N_A \times N_A$ random matrix $\boldsymbol{A}$, i.e., $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$, where each entry of $\boldsymbol{A}$ obeys the uniform distribution [0,1]. In this way, we obtain $\widetilde{\boldsymbol{W}}$. According to (5), we can obtain $\widetilde{\boldsymbol{W}}^{t_2}$, $t_2 = 1, 2, \ldots, T_2$, which is essentially dividing $\widetilde{\boldsymbol{W}}$ into T_2 submatrices. Then (14) is converted into T_2 subproblems, where each subproblem can be expressed as

$$\begin{aligned} \min_{\boldsymbol{W}_B^{t_2}, \boldsymbol{W}_R^{t_2}} \quad & \|\boldsymbol{W}_B^{t_2}\boldsymbol{W}_R^{t_2} - \widetilde{\boldsymbol{W}}^{t_2}\|_F \\ \textbf{s.t.} \quad & \boldsymbol{W}_R^{t_2} \in \mathcal{W}_{\mathcal{R}}, \|\boldsymbol{W}_B^{t_2}\boldsymbol{W}_R^{t_2}\|_F^2 = P_W. \end{aligned} \tag{15}$$

We cannot directly obtain solutions for (15) due to the non-convexity of the constraints. Note that the second constraint of (15) can be temporarily neglected during the optimization of $\boldsymbol{W}_B^{t_2}$ and $\boldsymbol{W}_R^{t_2}$. After $\boldsymbol{W}_B^{t_2}$ and $\boldsymbol{W}_R^{t_2}$ are obtained, we may set $\boldsymbol{W}_B^{t_2}$ as

$$\boldsymbol{W}_B^{t_2} \leftarrow \sqrt{P_W}\boldsymbol{W}_B^{t_2}/\|\boldsymbol{W}_B^{t_2}\boldsymbol{W}_R^{t_2}\|_F,\ t_2 = 1, 2, \ldots, T_2 \tag{16}$$

to satisfy the second constraint of (15).

To mitigate the interference among different data streams, we impose a common constraint that the columns of the digital precoding matrix are mutually orthogonal, i.e., ${\boldsymbol{W}_B^{t_2}}^H\boldsymbol{W}_B^{t_2} = \beta\boldsymbol{I}_{N_R}$, where $\beta \triangleq P_W/(N_A N_R)$. Define $\boldsymbol{W}_D^{t_2} \triangleq \beta^{-1}\boldsymbol{W}_B^{t_2}$. Then we have

$$\|\boldsymbol{W}_B^{t_2}\boldsymbol{W}_R^{t_2} - \widetilde{\boldsymbol{W}}^{t_2}\|_F^2 = \beta\|\boldsymbol{W}_R^{t_2} - {\boldsymbol{W}_D^{t_2}}^H\widetilde{\boldsymbol{W}}^{t_2}\|_F^2. \tag{17}$$

Therefore (15) can be expressed as

$$\min_{\boldsymbol{W}_D^{t_2}, \boldsymbol{W}_R^{t_2}} \|\boldsymbol{W}_R^{t_2} - {\boldsymbol{W}_D^{t_2}}^H\widetilde{\boldsymbol{W}}^{t_2}\|_F^2 \ \ \textbf{s.t.}\ \boldsymbol{W}_R^{t_2} \in \mathcal{W}_{\mathcal{R}}. \tag{18}$$

Algorithm 1 Hybrid Combining Matrix Design for IA-based Channel Estimation

1: *Input:* $\boldsymbol{D}(N_A, K)$, $\mathcal{W}_{\mathcal{R}}$, P_W, δ, $\widetilde{\boldsymbol{W}}$.
2: Obtain $\widetilde{\boldsymbol{W}}^{t_2}$ based on $\widetilde{\boldsymbol{W}}$ via (5), $t_2 = 1, 2, \ldots, T_2$.
3: **for** $t_2 = 1, 2, \ldots, T_2$ **do**
4: Set $i \leftarrow 0$, and obtain $\boldsymbol{W}_R^{t_2,i}$ randomly from $\mathcal{W}_{\mathcal{R}}$.
5: **repeat**
6: $i \leftarrow i + 1$.
7: Fix $\boldsymbol{W}_R^{t_2,i-1}$ and obtain $\boldsymbol{W}_B^{t_2,i}$ via (21).
8: Fix $\boldsymbol{W}_B^{t_2,i}$ and obtain $\boldsymbol{W}_R^{t_2,i}$ via (19).
9: **until** $\epsilon < \delta$
10: Update $\boldsymbol{W}_B^{t_2}$ via (16).
11: Obtain $\boldsymbol{W}^{t_2}$ via (4).
12: **end for**
13: Obtain $\boldsymbol{W}$ based on $\boldsymbol{W}^{t_2},\ t_2 = 1, 2, \ldots, T_2$ via (5).
14: *Output:* $\boldsymbol{W}$.

It shows that $\boldsymbol{W}_R^{t_2}$ and $\boldsymbol{W}_D^{t_2}$ are decoupled. Given $\boldsymbol{W}_D^{t_2}$, the solution of $\boldsymbol{W}_R^{t_2}$ can be expressed as

$$\boldsymbol{W}_R^{t_2} = \arg({\boldsymbol{W}_D^{t_2}}^H \widetilde{\boldsymbol{W}}^{t_2}, \mathcal{W}_{\mathcal{R}}) \tag{19}$$

where $\arg(\boldsymbol{A}, \mathcal{R})$ first normalizes each entry of $\boldsymbol{A}$ and then quantizes the normalized matrix in terms of $\mathcal{R}$. Note that the quantization is required since the resolution of phase shifters is limited in practice, e.g., the resolution is 2/64 if the phase shifter is 6 bits and the range of the angle is [-1,1].

Similarly, given $\boldsymbol{W}_R^{t_2}$, the optimization of $\boldsymbol{W}_D^{t_2}$ based on (18) can be expressed as

$$\begin{aligned} \min_{\boldsymbol{W}_D^{t_2}} \quad & \|\boldsymbol{W}_R^{t_2} - {\boldsymbol{W}_D^{t_2}}^H \widetilde{\boldsymbol{W}}^{t_2}\|_F^2 \\ \text{s.t.} \quad & {\boldsymbol{W}_D^{t_2}}^H \boldsymbol{W}_D^{t_2} = \beta^{-1} \boldsymbol{I}_{N_R}. \end{aligned} \tag{20}$$

It is seen that (20) is similar to the orthogonal Procrustes problem [10]. Then the solution to (20) can be obtained as

$$\boldsymbol{W}_D^{t_2} = \beta^{-1/2} \boldsymbol{V}\boldsymbol{U}^H, \tag{21}$$

where $\boldsymbol{W}_R^{t_2}(\widetilde{\boldsymbol{W}}^{t_2})^H = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$ represents the SVD of $\boldsymbol{W}_R^{t_2}(\widetilde{\boldsymbol{W}}^{t_2})^H$. Then we obtain $\boldsymbol{W}_B^{t_2} = \beta \boldsymbol{W}_D^{t_2}$.

As shown in **Algorithm 1**, we propose an algorithm of hybrid combining matrix design for IA-based channel estimation. We repeatedly fix $\boldsymbol{W}_R^{t_2}$ to obtain $\boldsymbol{W}_B^{t_2}$ via (21), and then fix $\boldsymbol{W}_B^{t_2}$ to obtain $\boldsymbol{W}_R^{t_2}$ via (19) in turn. Define the normalized iteration error ϵ as

$$\epsilon \triangleq \frac{\|\boldsymbol{W}_R^{t_2,i} - \boldsymbol{W}_R^{t_2,i-1}\|_F^2 + \|\boldsymbol{W}_B^{t_2,i} - \boldsymbol{W}_B^{t_2,i-1}\|_F^2}{\|\boldsymbol{W}_R^{t_2,i-1}\|_F^2 + \|\boldsymbol{W}_B^{t_2,i-1}\|_F^2}, \tag{22}$$

the stop condition is that the iterative update of both $\boldsymbol{W}_R^{t_2}$ and $\boldsymbol{W}_B^{t_2}$ is stable, i.e., $\epsilon < \delta$, where δ is the threshold.

The optimization problem for **hybrid precoding matrix design** is

$$\begin{aligned} \min_{\boldsymbol{F}_u} \quad & \|\boldsymbol{F}_u \boldsymbol{F}_u^H \boldsymbol{D}(M_A, K) - \gamma_M \boldsymbol{D}(M_A, K)\|_F \\ \text{s.t.} \quad & \boldsymbol{F}_{R,u}^{t_1} \in \mathcal{F}_{\mathcal{R}}, \|\boldsymbol{F}_{R,u}^{t_1} \boldsymbol{F}_{B,u}^{t_1}\|_F^2 = P_F,\ t_1 = 1, 2, \ldots, T_1, \end{aligned} \tag{23}$$

where $\mathcal{F}_{\mathcal{R}}$ is the set of all feasible analog precoding matrix and P_F is the given power for the hybrid precoding. Similar to (14), (23) can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{F}_u} \quad & \|\boldsymbol{F}_u \boldsymbol{F}_u^H - \gamma_M \boldsymbol{I}_{M_A}\|_F \\ \text{s.t.} \quad & \boldsymbol{F}_{R,u}^{t_1} \in \mathcal{F}_{\mathcal{R}}, \|\boldsymbol{F}_{R,u}^{t_1} \boldsymbol{F}_{B,u}^{t_1}\|_F^2 = P_F,\ t_1 = 1, 2, \ldots, T_1. \end{aligned} \tag{24}$$

We can obtain a solution as $\widetilde{\boldsymbol{F}}_u$ by selecting the first T_1 columns of $\sqrt{\gamma_M}\boldsymbol{U}$, where $\boldsymbol{U}$ is any $M_A \times M_A$ unitary matrix. According to (6), we can obtain $\widetilde{\boldsymbol{f}}_u^{t_1}$ as the t_1-th column of $\widetilde{\boldsymbol{F}}_u$, $t_1 = 1, 2, \ldots, T_1$. Define $\boldsymbol{f}_{B,u}^{t_1} \triangleq \boldsymbol{F}_{B,u}^{t_1} \boldsymbol{1}_{M_R},\ t_1 = 1, 2, \ldots, T_1$. Then we have $\boldsymbol{f}_u^{t_1} = \boldsymbol{F}_{R,u}^{t_1} \boldsymbol{f}_{B,u}^{t_1}$.

Similar to (15), (24) can be converted to T_1 subproblems, where each subproblem is expressed as

$$\begin{aligned} \min_{\boldsymbol{F}_{R,u}^{t_1},\ \boldsymbol{f}_{B,u}^{t_1}} \quad & \|\boldsymbol{F}_{R,u}^{t_1} \boldsymbol{f}_{B,u}^{t_1} - \widetilde{\boldsymbol{f}}_u^{t_1}\|_F \\ \text{s.t.} \quad & \boldsymbol{F}_{R,u}^{t_1} \in \mathcal{F}_{\mathcal{R}}, \|\boldsymbol{F}_{R,u}^{t_1} \boldsymbol{F}_{B,u}^{t_1}\|_F^2 = P_F. \end{aligned} \tag{25}$$

Similar to (15), the second constraint of (25) can also be temporarily neglected. Therefore, we may replace $\boldsymbol{W}_B^{t_2}$, $\boldsymbol{W}_R^{t_2}$, $\widetilde{\boldsymbol{W}}^{t_2}$, $\mathcal{W}_{\mathcal{R}}$ and P_W in (15) with $(\boldsymbol{f}_{B,u}^{t_1})^H$, $(\boldsymbol{F}_{R,u}^{t_1})^H$, $(\widetilde{\boldsymbol{f}}_u^{t_1})^H$, $\mathcal{F}_{\mathcal{R}}$ and P_F, respectively.

In order to run **Algorithm 1** to obtain $\boldsymbol{F}_u$, we have to further replace N_A, T_2, $\widetilde{\boldsymbol{W}}$ and $\boldsymbol{W}$ with M_A, T_1, $\widetilde{\boldsymbol{F}}_u$ and $\boldsymbol{F}_u$. The routine of the algorithm is exactly the same, except that an additional operation to obtain $\boldsymbol{F}_{B,u}^{t_1}$ as $\boldsymbol{F}_{B,u}^{t_1} = \frac{1}{M_R} \boldsymbol{f}_{B,u}^{t_1} \boldsymbol{1}_{M_R}^T$ is required after finishing step 9; and $\boldsymbol{W}_B^{t_2}$ should be replaced by $(\boldsymbol{F}_{B,u}^{t_1})^H$ at step 10.

In summary, in the first half of the IA-based channel estimation, we design of hybrid combining matrix and hybrid precoding matrix; while in the other half to be discussed, we will search the largest entry of the over-sampled beamspace channel matrix in an effective way.

C. Search the Largest Entry

It is time-consuming to exhaustively search the largest entry from $\boldsymbol{R}_u^v$ defined in (10). In order to improve the efficiency of the search algorithm, we analyze the structure of $\boldsymbol{R}_u^v$. Neglecting the term from the additive noise and assuming there is single path, we rewrite $\boldsymbol{R}_u^v$ as

$$\begin{aligned} \boldsymbol{R}_u^v &= \boldsymbol{D}(N_A, K)^H \boldsymbol{W}^H \boldsymbol{W} \boldsymbol{H}_u \boldsymbol{F}_u \boldsymbol{F}_u^H \boldsymbol{D}(M_A, K) \\ &= \sqrt{N_A M_A} g_{u,i} \boldsymbol{r}_N \boldsymbol{r}_M^H \end{aligned} \tag{26}$$

where $\boldsymbol{r}_N \triangleq \boldsymbol{D}(N_A, K)^H \boldsymbol{W}^H \boldsymbol{W} \boldsymbol{\alpha}(N_A, \theta_{u,1})$, $\boldsymbol{r}_M^H \triangleq \boldsymbol{\alpha}^H(M_A, \varphi_{u,1}) \boldsymbol{F}_u \boldsymbol{F}_u^H \boldsymbol{D}(M_A, K)$. Since $\boldsymbol{r}_N$ is a column vector and $\boldsymbol{r}_M^H$ is a row vector, the largest entry of $\boldsymbol{R}_u^v$ essentially depends on the largest entry of $\boldsymbol{r}_N$ and $\boldsymbol{r}_M$. Therefore we will analyze the structure of $\boldsymbol{r}_N$ and $\boldsymbol{r}_M$.

In the ideal case, i.e., $\boldsymbol{F}_u \boldsymbol{F}_u^H \boldsymbol{D}(M_A, K) = \gamma_M \boldsymbol{D}(M_A, K)$, $\boldsymbol{r}_M$ is an over-sampled transmit steering vector of $\boldsymbol{\alpha}(M_A, \varphi_{u,1})$ with an interval of $2/K$. In order to apply fast algorithms such as bisection search to find the peak of the curve, we have to first find the main lobe with the width of $4/M_A$; otherwise these algorithms may stop the search at the peak of side lobes.

Algorithm 2 Searching the Largest Entry Corresponding to the AoA and AoD of LOS Path for IA-based Channel Estimation

1: *Input:* $\boldsymbol{R}_u$.
2: **(First Stage)**
3: Obtain $\bar{\boldsymbol{R}}_u^v$ via (27).
4: Obtain s_q and s_p via (28) and (29), respectively.
5: Obtain $\boldsymbol{\Gamma} = [\Gamma_1, \Gamma_2]$ and $\boldsymbol{\Upsilon} = [\Upsilon_1, \Upsilon_2]$ via (30).
6: **(Second Stage)**
7: Obtain $\boldsymbol{R}_u^v$ via (10).
8: **while** $\Gamma_2 - \Gamma_1 > 2/K$ or $\Upsilon_2 - \Upsilon_1 > 2/K$ **do**
9:     Obtain Q_1, Q_2, Q_3 and Q_4 via (31).
10:     Obtain Q_{max} via (32).
11:     Update $\boldsymbol{\Gamma}$ and $\boldsymbol{\Upsilon}$.
12: **end while**
13: $\theta_{u,1} = \Gamma_2$, $\varphi_{u,1} = \Upsilon_2$
14: *Output:* $\theta_{u,1}$, $\varphi_{u,1}$.

Now we propose **Algorithm 2** to fast search the largest entry of $\boldsymbol{R}_u^v$, considering the structure of the steering vectors. **Algorithm 2** includes two stages. In the **first stage**, we find the main lobe of $\boldsymbol{r}_M$ and $\boldsymbol{r}_N$ without oversampling. In the **second stage** with oversampling, we apply the bisection search to find the peak of the main lobe.

In the **first stage** from step 3 to step 5, we search the main lobe, which is formed by two adjacent columns and two adjacent rows of $\bar{\boldsymbol{H}}_u^v$ defined in (8) [11]. We obtain the beamspace receiving matrix $\bar{\boldsymbol{R}}_u^v \in \mathbb{C}^{N_A \times M_A}$ at step 3 as

$$\bar{\boldsymbol{R}}_u^v = \boldsymbol{D}(N_A, N_A)^H \boldsymbol{W}^H \boldsymbol{W} \boldsymbol{H}_u \boldsymbol{F}_u \boldsymbol{F}_u^H \boldsymbol{D}(M_A, M_A) + \bar{\boldsymbol{n}}^v \tag{27}$$

where $\bar{\boldsymbol{n}}^v \triangleq \boldsymbol{D}(N_A, N_A)^H \boldsymbol{W}^H \tilde{\boldsymbol{n}} \boldsymbol{F}_u^H \boldsymbol{D}(M_A, M_A)$ is a noise term. We first find two adjacent columns indexed by $\{s_q, s_q+1\}$ with the largest channel power from $\bar{\boldsymbol{H}}_u^v$ at step 4 via

$$s_q = \arg \max_{q=1,2,\ldots,M_A-1} \|\bar{\boldsymbol{R}}_{u,q}^v\|_F \tag{28}$$

where $\bar{\boldsymbol{R}}_{u,q}^v$ represents a submatrix consisted of two consecutive columns of $\bar{\boldsymbol{R}}_u^v$, with column indices denoted as q and $q+1$. Similarly, we find two adjacent rows indexed by $\{s_p, s_p+1\}$ with the largest channel power from $\bar{\boldsymbol{R}}_u^v$ at step 4 via

$$s_p = \arg \max_{p=1,2,\ldots,N_A-1} \|\bar{\boldsymbol{R}}_{u,p}^v\|_F \tag{29}$$

where $\bar{\boldsymbol{R}}_{u,p}^v$ represents a submatrix consisted of two consecutive rows of $\bar{\boldsymbol{R}}_u^v$, with row indices denoted as p and $p+1$. By finding out the largest two adjacent columns $\{s_q, s_q+1\}$ and rows $\{s_p, s_p+1\}$ of beamspace channel matrix $\bar{\boldsymbol{H}}_u^v$, the search of AoA and AoD can be limited to the range of

$$\boldsymbol{\Gamma} = [\Gamma_1, \Gamma_2], \;\; \boldsymbol{\Upsilon} = [\Upsilon_1, \Upsilon_2], \tag{30}$$

respectively, where $\Gamma_1 \triangleq -1 + 2(s_p - 3/2)/N_A$, $\Gamma_2 \triangleq -1 + 2(s_p + 1/2)/N_A$, $\Upsilon_1 \triangleq -1 + 2(s_q - 3/2)/M_A$, $\Upsilon_2 \triangleq -1 + 2(s_q + 1/2)/M_A$. In this way, we narrow down the search space of the AoA and AoD from $[-1, 1]$ to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Upsilon}$, respectively.

In the **second stage** from step 7 to step 13, we find the coordinates of the largest entry of $\boldsymbol{R}_u^v$ corresponding to the AoA in $\boldsymbol{\Gamma}$ and AoD in $\boldsymbol{\Upsilon}$. Note that the quantization error of the AoD and AoA is reduced from $2/M_A$ and $2/N_A$ to both $2/K$ by oversampling. We apply the bisection search. Define *quan*() as the quantization function to quantize the consecutive θ into K discrete points, which is denoted as $\mathit{quan}(\theta) \triangleq \left\langle \frac{K(\theta+1)}{2} \right\rangle$. The two points that divide $\boldsymbol{\Gamma}$ into two equal parts are $\mathit{quan}(\Gamma_1/2+\Gamma_2/2)$ and $\mathit{quan}(\Gamma_1/2+\Gamma_2/2)+1$. Similarly, the two points that divide $\boldsymbol{\Upsilon}$ into two equal parts are $\mathit{quan}(\Upsilon_1/2+\Upsilon_2/2)$ and $\mathit{quan}(\Upsilon_1/2+\Upsilon_2/2)+1$. The entries corresponding to these four points are

$$\begin{aligned} Q_1 &= \boldsymbol{R}_u^v[\mathit{quan}(\Gamma_1/2+\Gamma_2/2), \mathit{quan}(\Upsilon_1/2+\Upsilon_2/2)], \\ Q_2 &= \boldsymbol{R}_u^v[\mathit{quan}(\Gamma_1/2+\Gamma_2/2), \mathit{quan}(\Upsilon_1/2+\Upsilon_2/2)+1], \\ Q_3 &= \boldsymbol{R}_u^v[\mathit{quan}(\Gamma_1/2+\Gamma_2/2)+1, \mathit{quan}(\Upsilon_1/2+\Upsilon_2/2)], \\ Q_4 &= \boldsymbol{R}_u^v[\mathit{quan}(\Gamma_1/2+\Gamma_2/2)+1, \mathit{quan}(\Upsilon_1/2+\Upsilon_2/2)+1]. \end{aligned} \tag{31}$$

Then we compare the amplitude of Q_1, Q_2, Q_3 and Q_4 to find the largest one, which is expressed as

$$Q_{max} = \max\left\{|Q_1|, |Q_2|, |Q_3|, |Q_4|\right\}. \tag{32}$$

We delete the parts that do not include Q_{max} and update $\boldsymbol{\Gamma}$ and $\boldsymbol{\Upsilon}$. For example, if $Q_{max} = Q_1$, the updated Γ_1, Γ_2, Υ_1 and Υ_2, denoted as $\bar{\Gamma}_1$, $\bar{\Gamma}_2$, $\bar{\Upsilon}_1$ and $\bar{\Upsilon}_2$, respectively, can be represented as $\bar{\Gamma}_1 = \Gamma_1$, $\bar{\Gamma}_2 = \Gamma_1/2 + \Gamma_2/2$, $\bar{\Upsilon}_1 = \Upsilon_1$, $\bar{\Upsilon}_2 = \Upsilon_1/2 + \Upsilon_2/2$. We repeat the procedures until $\Gamma_2 - \Gamma_1 \le 2/K$ and $\Upsilon_2 - \Upsilon_1 \le 2/K$, which means the resolution $2/K$ of over-sampled search is reached for both the AoA and AoD. Finally we output the estimated AoA and AoD of LOS path at step 14.

D. Computational Complexity

Now we compare the computational complexity for the IA-based, DCS-based and OCS-based schemes.

Since the hybrid precoding matrix and hybrid combining matrix can be designed off-line before the channel training, the computational complexity mainly comes from the search of the largest entry in **Algorithm 2**. For IA-based scheme, in the first stage, we need to compute Frobenius norm in (28) and (29), resulting in the complexity to be $\mathcal{O}(4N_A(M_A-1) + 4M_A(N_A-1))$. In the second stage, we use the bisection search to find the largest entry among $2K/N_A \times 2K/M_A$ entries. In the first $\log_2(2K/N_A)$ iterations, we compute four bisection points in each iteration. In the following $\log_2(2K/M_A) - \log_2(2K/N_A)$ iterations, since the AoA has been estimated, we only need to compute two bisection points in each iteration. Therefore the total computational complexity for the IA-based scheme is

$$\begin{aligned} &\mathcal{O}(4N_A(M_A-1) + 4M_A(N_A-1) + 8\log_2(2K/N_A) \\ &(N_A+1)M_A + 4(\log_2(2K/M_A) - \log_2(2K/N_A))M_A). \end{aligned} \tag{33}$$

For the DCS-based scheme [4], where the main lobe is first searched and then the largest entry within the main lobe is further searched by exhaustive search, the computational complexity is $\mathcal{O}(2N_A(M_A-1) + 2M_A(N_A-1) + 2K^2(N_A+1)/N_A)$. For the OCS-based scheme [5], the largest entry

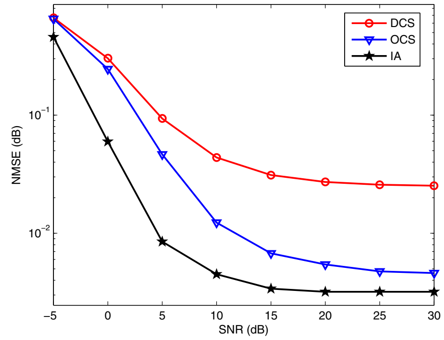

Fig. 1. Comparisons of channel estimation performance in terms of NMSE.

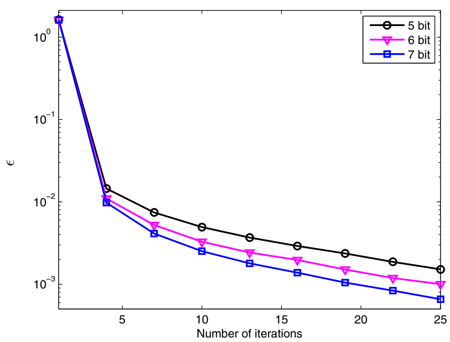

Fig. 2. Convergence of **Algorithm 1**.

is directly searched by exhaustive search, the computational complexity is $\mathcal{O}(2M_A(N_A+1)K^2)$. Since $K > N_A$ and $K > M_A$, the computational complexity of the IA-based scheme is much lower than that of DCS-based and OCS-based schemes.

IV. Simulation Results

We consider the uplink multiuser mmWave massive MIMO system, which includes a BS and $U = 4$ users. The BS is equipped with $N_A = 64$ antennas and $N_R = 4$ RF chains, while each user is equipped with $M_A = 16$ antennas and $M_R = 1$ RF chain. The number of resolvable paths in mmWave channel is set to be $L_u = 3$, while $g_{u,1} \sim \mathcal{CN}(0,1)$ and $g_{u,i} \sim \mathcal{CN}(0,0.01)$ for $i = 2,3$. We use $K = 1024$ over-sampled steering vectors.

As shown in Fig. 1, we compare the channel estimation performance in terms of normalized mean square error (N-MSE) for the proposed IA-based scheme, the DCS-based scheme [4] and the OCS-based scheme [5]. We set $T_1 = 4$ and $T_2 = 4$. Then $T_3 = T_2 N_R = 16$. In order to make fair comparisons, we set the total time slots of the DCS-based scheme and the OCS-based scheme the same as the IA-based scheme. The number of total time slots for pilot training is fairly set to be $UT_1T_2 = 64$. Suppose we use 6 bit and 4 bit digital phase shifters at the BS and users, respectively. It is observed from Fig. 1 that IA-based scheme outperforms the DCS-based and OCS-based schemes. At SNR of 15dB, IA-based scheme has 89.1% and 49.7% performance improvement compared with the DCS-based and OCS-based schemes, respectively. The reason is that both DCS-based and OCS-based schemes employ random precoding and random combining matrix, which are not optimal.

As shown in Fig. 2, we verify the convergence of **Algorithm 1**. Suppose we use 5, 6 and 7 bit digital phase shifters at the BS, respectively. It is seen that ϵ in (22) decreases rapidly as the number of iterations grows, and ϵ decreases faster with higher resolution of digital phase shifters. Using 6 bit digital phase shifters at the BS, ϵ is smaller than 10^{-3} when the number of iterations is larger than 25, which means 25 iterations is enough for $\delta = 10^{-3}$.

V. Conclusions

In this paper, we have proposed an IA-based channel estimation scheme, which includes the design of the hybrid precoding and combining matrix as well as searching the largest entry of over-sampled beamspace receiving matrix. Future work will focus on the interference mitigation for multiuser mmWave MIMO transmission.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61302097 and by the Natural Science Foundation of Jiangsu Province under Grant BK20161428.

References

[1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[2] A. Liao, Z. Gao, Y. Wu, H. Wang, and M.-S. Alouini, "2D unitary ESPRIT based super-resolution channel estimation for millimeter-wave massive MIMO with hybrid precoding," *IEEE Access*, vol. 5, pp. 24 747–24 757, Nov. 2017.

[3] X. Gao, L. Dai, S. Han, C.-L. I, and F. Adachi, "Beamspace channel estimation for 3D lens-based millimeter-wave massive MIMO systems," in *Proc. IEEE WCSP*, Yangzhou, China, Oct. 2016, pp. 1–5.

[4] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

[5] A. Alkhateeb, G. Leus, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. IEEE ICASSP*, Dubai, UAE, Apr. 2015, pp. 2909–2913.

[6] P. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2222, Jun. 2015.

[7] Y. Sun and C. Qi, "Weighted sum-rate maximization for analog beamforming and combining in millimeter wave massive MIMO communications," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1883–1886, Jun. 2017.

[8] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, 2017.

[9] A. Jennings and J. J. McKeown, *Matrix computation*. Wiley, 1992.

[10] J. C. Gower and G. B. Dijksterhuis, *Procrustes problems*. London, U.K.: Oxford Univ. Press, 2004.

[11] W. Ma and C. Qi, "Channel estimation for 3-D lens millimeter wave massive MIMO system," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 2045–2048, Jun. 2017.