Super-Resolution Channel Estimation for MmWave Massive MIMO with Hybrid Precoding

Chen Hu, Linglong Dai, Senior Member, IEEE, Zhen Gao, Member, IEEE, and Jun Fang, Senior Member, IEEE

Abstract—Channel estimation is challenging for millimeter-wave (mmWave) massive MIMO with hybrid precoding, since the number of radio frequency (RF) chains is much smaller than that of antennas. Conventional compressive sensing based channel estimation schemes suffer from severe resolution loss due to the channel angle quantization. To improve the channel estimation accuracy, we propose an iterative reweight (IR)-based super-resolution channel estimation scheme in this paper. By optimizing an objective function through the gradient descent method, the proposed scheme can iteratively move the estimated angle of arrivals/departures (AoAs/AoDs) towards the optimal solutions, and finally realize the super-resolution channel estimation. In the optimization, a weight parameter is used to control the tradeoff between the sparsity and the data fitting error. In addition, a singular value decomposition (SVD)-based preconditioning is developed to reduce the computational complexity of the proposed scheme. Simulation results verify the better performance of the proposed scheme than conventional solutions.

Index Terms—Millimeter-wave (mmWave), massive MIMO, hybrid precoding, angle of arrival (AoA), angle of departure (AoD), super-resolution channel estimation.

I. INTRODUCTION

Millimeter-wave (mmWave) massive MIMO has been recognized as a promising technology for future 5G wireless communications [1]. To reduce the hardware cost and power consumption, hybrid precoding has been proposed for practical mmWave massive MIMO systems, where hundreds of antennas are driven by a much smaller number of radio frequency (RF) chains [2], [3]. The analog and digital co-design in hybrid precoding requires accurate channel state information. However, the digital baseband cannot directly access all antennas due to the small number of RF chains, so it is difficult to accurately estimate the high-dimensional MIMO channel [4], [5].

Several novel channel estimation schemes have been recently proposed for mmWave massive MIMO with hybrid precoding [5]–[9]. Specifically, [5], [6] proposed the adaptive codebook-based channel sounding scheme, where the transmitter and receiver search for the best beam pair by adjusting the predefined precoding and combining codebooks. However, the channel estimation resolution is limited by the codebook size. [7] was able to achieve better angle estimation by performing an amplitude comparison with respect to the auxiliary beam pair. On the other hand, by exploiting the angular channel sparsity, the on-grid compressive sensing based methods [8], [9] could estimate the channel with reduced training overhead. However, such solutions assumed that the angle of arrivals/departures (AoAs/AoDs) lie in discrete points in the angle domain (i.e., “on-grid” AoAs/AoDs), while the actual AoAs/AoDs are continuously distributed (i.e., “off-grid” AoAs/AoDs) in practice. The assumption of on-grid AoAs/AoDs result in the power leakage problem, which severely degrades the channel estimation accuracy.

In order to improve the channel estimation accuracy, we propose an iterative reweight (IR)-based super-resolution channel estimation scheme in this paper. Inspired by the IR approach recently proposed for 1-dimensional line spectral estimation [10], we extend this approach to 2-dimensional mmWave channel estimation. Specifically, we iteratively optimize the estimates of AoAs/AoDs, to decrease the weighted summation of the sparsity and the data fitting error. The weight controlling the tradeoff between the sparsity and the data fitting error, is iteratively updated to avoid over-fitting or under-fitting. Since the estimated AoAs/AoDs can be moved from the initial angle-domain grids towards the actual off-grid AoAs/AoDs, the proposed scheme is able to achieve super-resolution channel estimation. In addition, we propose a singular value decomposition (SVD)-based preconditioning method to reduce the computational complexity of the proposed scheme, which is realized by reducing the number of initial candidates of AoAs/AoDs in the IR procedure. Simulation results show that the proposed IR-based super-resolution channel estimation can achieve better performance than conventional solutions.

Notation: In this paper, the boldface lower and upper-case symbols denote column vectors and matrices, respectively. $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ denote the transpose, the conjugate transpose, and the inverse of a matrix, respectively. $\text{diag}(x)$ is the diagonal matrix with the vector $x$ on its diagonal. The $\ell_0$-norm, $\ell_2$-norm, and Frobenius norm are given by $\| \cdot \|_0$, $\| \cdot \|_2$, and $\| \cdot \|_F$, respectively.

II. SYSTEM MODEL

We consider a typical mmWave massive MIMO system with uniform linear arrays (ULAs). Let $N_T$, $N_R$, $N_T^{\text{RF}}$, and $N_R^{\text{RF}}$ be the number of transmit antennas, receive antennas, transmitter RF chains, and receiver RF chains, respectively. For practical mmWave massive MIMO with hybrid precoding,
the number of RF chains is much smaller than that of antennas, i.e., \(N_{\text{RF}} < N_{\text{T}}\), \(N_{\text{RF}} < N_{\text{R}}\) [1]–[3]. The system model can be given by

\[
r = QH\Phi P + n,
\]

where \(r \in \mathbb{C}^{N_{\text{RF}} \times 1}\) is the received signal in the baseband, \(Q \in \mathbb{C}^{N_{\text{RF}} \times N_{\text{R}}}\) is the hybrid combining matrix, \(H \in \mathbb{C}^{N_{\text{R}} \times N_{\text{T}}}\) is the channel matrix, \(P \in \mathbb{C}^{N_{\text{T}} \times N_{\text{RF}}}\) is the hybrid precoding matrix, \(s \in \mathbb{C}^{N_{\text{RF}} \times 1}\) is the transmitted baseband signal, and \(n \in \mathbb{C}^{N_{\text{RF}} \times 1}\) is the received noise after combining.

The channel model

\[
H = \sum_{l=1}^{L} z_l \alpha_l (d \sin \phi_{l1}/\lambda) a_l^H (d \sin \phi_{lT}/\lambda) \tag{2}
\]

is widely adopted in millimeter-wave MIMO systems [5]–[9], where \(L \ll \min(N_{\text{R}}, N_{\text{T}})\) is the number of propagation paths, \(d\) is the antenna spacing, \(\lambda\) is the wavelength, \(z_l, \phi_{l1}\), and \(\phi_{lT}\) are the complex path gain, the physical AoA and AoD of the \(l\)-th path, respectively. \(\alpha_l (d \sin \phi_{l1}/\lambda)\) and \(a_l (d \sin \phi_{lT}/\lambda)\) are the steering vector at the receiver and the steering vector at the transmitter of the \(l\)-th path, respectively. For typical ULAs, \(\alpha_l (d \sin \phi_{l1}/\lambda)\) and \(a_l (d \sin \phi_{lT}/\lambda)\) are given by

\[
\alpha_l (d \sin \phi_{l1}/\lambda) = \left[1, e^{j2\pi d \sin \phi_{l1}/\lambda}, \ldots, e^{j2\pi (N_{\text{R}}-1)d \sin \phi_{l1}/\lambda}\right]^T,
\]

\[
a_l (d \sin \phi_{lT}/\lambda) = \left[1, e^{j2\pi d \sin \phi_{lT}/\lambda}, \ldots, e^{j2\pi (N_{\text{T}}-1)d \sin \phi_{lT}/\lambda}\right]^T. \tag{3}
\]

By defining \(\theta_{l1} \Delta = d \sin \phi_{l1}/\lambda\) and \(\theta_{lT} \Delta = d \sin \phi_{lT}/\lambda\) as the normalized spacial angles, the mmWave channel matrix \(H\) in (2) can be also written as

\[
H = A_l (\theta_{l1}) \text{diag}(z) A_l^H (\theta_{lT}), \tag{4}
\]

where \(z = [z_1, z_2, \ldots, z_L]^T\), \(\theta_{l1} = [\theta_{l11}, \theta_{l12}, \ldots, \theta_{l1L}]^T\), \(\theta_{lT} = [\theta_{lT1}, \theta_{lT2}, \ldots, \theta_{lT,L}]^T\), \(A_l (\theta_{l1}) = \left[\alpha_l(\theta_{l1}), \alpha_l(\theta_{l12}), \ldots, \alpha_l(\theta_{l1L})\right]^T\), \(A_l (\theta_{lT}) = \left[a_l(\theta_{lT1}), a_l(\theta_{lT2}), \ldots, a_l(\theta_{lT,L})\right]^T\).

Denote \(x = P s \in \mathbb{C}^{N_{\text{T}} \times 1}\), where the \(i\)-th element of \(x\) is the transmitted signal at the \(i\)-th transmit antenna. Suppose that the transmitter sends \(N_{\text{T}}\) different pilot sequences \(x_1, x_2, \ldots, x_{N_{\text{T}}}\). Since the number of RF chains is smaller than the required dimension of received pilot sequence, for each transmit pilot sequence \(x_p\) (\(1 \leq p \leq N_{\text{T}}\)), we use \(M\) time slots to obtain an \(N_{\text{R}}\)-dimensional received pilot sequence \(y_p\), where \(N_{\text{R}} = MN_{\text{RF}}\). Thus, the training overhead is \(T = MN_{\text{T}}\). In the \(m\)-th time slot, we use the combining matrix \(W_m\) to obtain an \(N_{\text{R}}\)-dimensional received pilot sequence

\[
y_{p,m} = W_m^H H x_p + n_{p,m}. \tag{5}
\]

By collecting the received pilots in the \(M\) time slots, we have

\[
y_p = W_m^H H x_p + n_p, \quad W_p = [y_{p,1}, y_{p,2}, \ldots, y_{p,M}]^T \in \mathbb{C}^{N_{\text{R}} \times 1}, \quad W = [W_1, W_2, \ldots, W_M] \in \mathbb{C}^{N_{\text{R}} \times N_{\text{R}}}, \quad n_p \in \mathbb{C}^{N_{\text{R}} \times 1}
\]

The estimation of the channel matrix \(H\) in (6) is equivalent to the estimation of the normalized spacial angles \((\theta_{l1}, \theta_{lT})\) and path gains \(z_l\) for all \(L\) paths \((l = 1, 2, \ldots, L)\). However, the number of paths \(L\) is not \textit{a priori} in practice. Due to the angle-domain sparsity of the channel matrix \(H\) [1], the sparse channel estimation problem can be formulated as

\[
\min_{\hat{z}, \hat{\theta}_{l1}, \hat{\theta}_{lT}} \|\hat{z}\|_0, \quad \text{s.t.} \quad \|Y - W^H \hat{H} X\|_F \leq \varepsilon, \tag{7}
\]

where \(\|\hat{z}\|_0\) is the number of non-zero elements of \(\hat{z}\), which means the estimated number of paths \(\hat{L}\), \(\hat{H}\) is the estimated channel matrix, and \(\varepsilon\) is the error tolerance parameter [9].

III. PROPOSED IR-BASED SUPER-RESOLUTION CHANNEL ESTIMATION

In this section, we firstly extend the 1-dimensional iterative reweighted (IR) method for line spectral estimation [10] to the 2-dimensional form for channel estimation. Then, we optimize the angle estimates via gradient descent method to achieve super-resolution channel estimation. Finally, we propose a singular value decomposition (SVD)-based preconditioning to reduce the computational complexity.

A. Proposed Optimization Formulation

The main difficulty in solving (7) lies in the fact that the \(L_0\)-norm is not computationally efficient for finding the optimal solution. By replacing the \(L_0\)-norm with a log-sum function [10], we have

\[
\min_{\hat{z}, \hat{\theta}_{l1}, \hat{\theta}_{lT}} F(\hat{z}) \Delta \sum_{l=0}^{L} \log \left(\|z_l\|^2 + \delta\right), \quad \text{s.t.} \quad \|Y - W^H \hat{H} X\|_F \leq \varepsilon, \tag{8}
\]

where \(\delta > 0\) ensures that the logarithmic function is well-defined [10]. \(\hat{H}\) is determined by the parameters \(z, \theta_{l1}, \theta_{lT}\) defined in (4). By adding a regularization parameter \(\lambda > 0\), we can further formulate the problem (8) as an unconstrained optimization problem:

\[
\min_{\hat{z}, \hat{\theta}_{l1}, \hat{\theta}_{lT}} G(\hat{z}, \hat{\theta}_{l1}, \hat{\theta}_{lT}) \Delta \sum_{l=1}^{L} \log \left(\|z_l\|^2 + \delta\right) + \lambda \|Y - W^H \hat{H} X\|_F^2, \tag{9}
\]

Moreover, by using an iterative surrogate function instead of the log-sum function, the minimization of \(G(\hat{z}, \hat{\theta}_{l1}, \hat{\theta}_{lT})\) is equivalent to the minimization of the surrogate function [10]:

\[
\min_{\hat{z}, \hat{\theta}_{l1}, \hat{\theta}_{lT}} S^{(i)}(\hat{z}, \hat{\theta}_{l1}, \hat{\theta}_{lT}) \Delta -\lambda - 1 z^H D^{(i)} z + \|Y - W^H \hat{H} X\|_F^2, \tag{10}
\]

where \(D^{(i)}\) is defined as

\[
D^{(i)} \Delta \text{diag} \left(\frac{1}{\|z_{1}^{(i)}\|^2 + \delta}, \frac{1}{\|z_{2}^{(i)}\|^2 + \delta}, \ldots, \frac{1}{\|z_{L}^{(i)}\|^2 + \delta}\right). \tag{11}
\]

where \(z_{i}^{(i)}\) is the estimate of \(z\) at the \(i\)-th iteration. The reason why the minimization of the log-sum function \(G(\hat{z}, \hat{\theta}_{l1}, \hat{\theta}_{lT})\) in (9) can be equivalently transformed to the minimization of \(S^{(i)}(\hat{z}, \hat{\theta}_{l1}, \hat{\theta}_{lT})\) in (10) can be explained as follows. In the \(i\)th iteration, better estimates \(\hat{z}^{(i+1)}, \hat{\theta}_{l1}^{(i+1)}, \hat{\theta}_{lT}^{(i+1)}\) can be found to make \(S^{(i)}\) become smaller, i.e.,

\[
S^{(i)} \left(\hat{z}^{(i+1)}, \hat{\theta}_{l1}^{(i+1)}, \hat{\theta}_{lT}^{(i+1)}\right) \leq S^{(i)} \left(\hat{z}^{(i)}, \hat{\theta}_{l1}^{(i)}, \hat{\theta}_{lT}^{(i)}\right). \tag{12}
\]
Then, we have
\[
G \left( \hat{z}^{(i+1)}, \hat{\theta}_R^{(i+1)}, \hat{\theta}_T^{(i+1)} \right) - \lambda S^{(i)} \left( \hat{z}^{(i+1)}, \hat{\theta}_R^{(i+1)}, \hat{\theta}_T^{(i+1)} \right)
\]
\[
= \sum_{l=1}^{L} \left[ \log \left( \frac{\hat{z}_l^{(i+1)}}{\hat{z}_l^{(i)}} \right)^2 + \delta \right] - \left[ \frac{\hat{z}_l^{(i+1)}}{\hat{z}_l^{(i)}} \right]^2 + \delta
\]
\[
\leq \sum_{l=1}^{L} \left[ \log \left( \frac{\hat{z}_l^{(i)}}{\hat{z}_l^{(i)}} \right)^2 + \delta \right] - \left[ \frac{\hat{z}_l^{(i+1)}}{\hat{z}_l^{(i)}} \right]^2 + \delta
\]
\[
= G \left( \hat{z}^{(i)}, \hat{\theta}_R^{(i)}, \hat{\theta}_T^{(i)} \right) - \lambda S^{(i)} \left( \hat{z}^{(i)}, \hat{\theta}_R^{(i)}, \hat{\theta}_T^{(i)} \right).
\]
(13)

This is because \( G - \lambda S^{(i)} \) reaches its maximum value at \( \hat{z}^{(i+1)} = \hat{z}^{(i)} \). With (12) and (13), we have
\[
G \left( \hat{z}^{(i+1)}, \hat{\theta}_R^{(i+1)}, \hat{\theta}_T^{(i+1)} \right) \leq G \left( \hat{z}^{(i)}, \hat{\theta}_R^{(i)}, \hat{\theta}_T^{(i)} \right),
\]
which indicates that the minimization of \( G \) is equivalent to the minimization of \( S^{(i)} \).

Then, as proved in Appendix A, we can optimize (10) with regard to the path gains \( z \), to find the optimal point of \( \hat{z} \) and the corresponding optimal value of \( S^{(i)} \) as follows:
\[
\hat{z}^{(i)}(\theta_R, \theta_T) \triangleq \arg \min_z S^{(i)}(z, \theta_R, \theta_T)
\]
\[
= \left( \lambda^{-1} D^{(i)} + \sum_{p=1}^{N_x} K_p^H K_p \right)^{-1} \left( \sum_{p=1}^{N_x} K_p^H Y_p \right),
\]
(15)
\[
S^{(i)}(\theta_R, \theta_T) \triangleq \min_z S^{(i)}(z, \theta_R, \theta_T)
\]
\[
= \left( \sum_{p=1}^{N_x} Y_p^H K_p \right) \left( \lambda^{-1} D^{(i)} + \sum_{p=1}^{N_x} K_p^H K_p \right)^{-1} \left( \sum_{p=1}^{N_x} K_p^H Y_p \right) + \sum_{p=1}^{N_x} Y_p^H Y_p,
\]
(16)
where \( K_p = W^H A_R \text{diag} \left( A_p^H \theta_p \right) \). After that, we only need to optimize the normalized spatial angles \( \theta_R \) and \( \theta_T \) in (16), which will be discussed in the next subsection.

B. IR-Based Super-Resolution Channel Estimation

In the previous subsection, we have already simplified the constrained optimization problem (7) to an unconstrained angle optimization problem (16). To solve this reformulated problem, now we propose an IR-based super-resolution channel estimation scheme as described in Algorithm 1.

The objective function \( S^{(i)}(z, \theta_R, \theta_T) \) is the weighted sum of two parts: \( z^H D z \) controlling the sparsity of the estimation result and \( \| Y - W^H H X \|_F \) denoting the residue. In addition, \( \lambda \) is the regularization parameter that controls the tradeoff between the sparsity and the data fitting error. A larger \( \lambda \) implies a well-fitting solution, while a smaller \( \lambda \) means a more sparse but worse-fitting solution. Thus, the choice of \( \lambda \) is critical to the channel estimation performance and the convergence speed. In the iterative reweighted method [10], \( \lambda \) is not fixed but updated in each iteration. To be specific, if the previous iteration is poorly-fitted, we will choose a smaller \( \lambda \) to make the estimate sparser. On the other hand, if the previous iteration returns a well-fitted estimate and leads to a small residue, our method will choose a larger \( \lambda \) to accelerate the searching for the best-fitting estimate. In the proposed algorithm, \( \lambda \) is updated by
\[
\lambda = \min \left\{ d/r^{(i)}, \lambda_{\text{max}} \right\},
\]
(17)
where \( d \) is a constant scaling factor, and \( \lambda_{\text{max}} \) is selected to make the problem well-conditioned, \( r^{(i)} \) is the squared residue in the previous step, i.e.,
\[
r^{(i)} = \left\| Y - W^H A_R \left( \hat{\theta}_R^{(i)}, \hat{\theta}_T^{(i)} \right) \text{diag}(\hat{z}^{(i)}) A_R^H \hat{\theta}_T^{(i)} X \right\|_F^2.
\]
(18)

The proposed algorithm starts iteration at the angle domain grids. In the \( i \)-th iteration, our task is to search for new estimates \( \hat{\theta}_R^{(i+1)} \) and \( \hat{\theta}_T^{(i+1)} \) in the neighborhood of the previous estimates \( \hat{\theta}_R^{(i)} \) and \( \hat{\theta}_T^{(i)} \) to make the objective function \( S^{(i)} \) become smaller. This searching can be accomplished via gradient descent method:
\[
\hat{\theta}_R^{(i+1)} = \hat{\theta}_R^{(i)} - \eta \cdot \nabla_{\theta_R} S^{(i)}(\hat{\theta}_R^{(i)}, \hat{\theta}_T^{(i)}),
\]
\[
\hat{\theta}_T^{(i+1)} = \hat{\theta}_T^{(i)} - \eta \cdot \nabla_{\theta_T} S^{(i)}(\hat{\theta}_R^{(i)}, \hat{\theta}_T^{(i)}),
\]
(19)
where the gradients can be calculated according to Appendix B, and \( \eta \) is the chosen step-length to make sure \( S^{(i)}(\hat{\theta}_R^{(i+1)}, \hat{\theta}_T^{(i+1)}) \leq S^{(i)}(\hat{\theta}_R^{(i)}, \hat{\theta}_T^{(i)}) \).

The estimates \( \hat{\theta}_R^{(i)} \) and \( \hat{\theta}_T^{(i)} \) become more and more accurate during the iterative searching, until the new estimates are almost the same as the previous ones. The final estimates of channel parameters \( \hat{\theta}_R, \hat{\theta}_T, \hat{z} \) converge to their true values with high probability, which will be verified later by simulations. Theoretical analysis of the IR algorithm has been recently provided in [10]. With our proposed IR-based super-resolution channel estimation scheme, the estimates of \( \hat{\theta}_R, \hat{\theta}_T \) can be moved from the initial on-grid coarse estimates to its actual off-grid positions, thus the super-resolution channel estimation can be realized.

It is worthy to point out that the sparsity level \( L \) is unknown in practice. In the proposed scheme, the sparsity level can be initialized to be larger than the real channel sparsity. During the iteration process, the paths with too small path gains will be regarded as noise instead of real paths. Then, our algorithm prune these paths to make the result sparser. By iteratively pruning these paths, the estimated sparsity level will decrease to the real number of paths.

Algorithm 1 IR-based super-resolution channel estimation

Input: Noisy received signals \( Y \), transmit pilot signals \( X \), combining matrix \( W \), initial on-grid AoAs and AoDs \( \hat{\theta}_R^{(0)}, \hat{\theta}_T^{(0)} \), pruning threshold \( \varepsilon_{z_R} \) and termination threshold \( \varepsilon_{\text{th}} \).

Output: Estimated AoAs/AoDs and path gains of all paths.

1: Initialize \( \hat{z}^{(0)} = z_{\text{opt}} \left( \hat{\theta}_R^{(0)}, \hat{\theta}_T^{(0)} \right) \) according to (15).
2: repeat
3: Update \( \lambda \) by (17).
4: Construct the function \( S^{(i)}_{\text{opt}}(\theta_R, \theta_T) \) by (16).
5: Search for new angle estimates \( \hat{\theta}_R^{(i+1)}, \hat{\theta}_T^{(i+1)} \) by (19).
6: Estimate the path gains \( \hat{z}^{(i+1)} \) according to (15).
7: Prune path \( l \) if \( \frac{\hat{z}_l^{(i+1)}}{\hat{z}_l^{(i)}} < \varepsilon_{z_R} \).
8: until \( L^{(i)} = L^{(i+1)} \) and \( \| \hat{z}^{(i+1)} - \hat{z}^{(i)} \|_2 < \varepsilon_{\text{th}} \).
9: \( \hat{\theta}_R = \hat{\theta}_R^{(\text{last})}, \hat{\theta}_T = \hat{\theta}_T^{(\text{last})}, \hat{z} = \hat{z}^{(\text{last})} \).
The computational complexity in each iteration lies in calculating the gradient in Step 5. It takes \( O\left(N_XN_Y(N_R + N_T)L\right) \) calculations for each partial derivatives against one parameter, so the computational complexity to calculate the gradient is \( O\left(N_XN_Y(N_R + N_T)^2L^2\right) \). As a result, the number of initial candidates \( L(0) \) is critical, and it should be as small as possible to make the computation affordable. The problem how to effectively select the initial \( \theta_R^{(0)} \) and \( \theta_T^{(0)} \) before the iteration will be discussed in the next section.

C. SVD-based Preconditioning

In this section, we propose a singular value decomposition (SVD)-based preconditioning as shown in Algorithm 2, to reduce the computational complexity of the IR procedure in the proposed IR-based super-resolution channel estimation scheme. The proposed scheme can find the angle-domain grids nearest to the real AoAs/AoDs by finding the best matching grids (steps 4-7).

Specifically, by applying SVD to the matrix \( Y \), we have

\[
Y = UV^T + N.
\]

As the noise is small, the largest \( L \) singular values and their corresponding singular vectors are approximately determined by the \( L \) paths, i.e., for \( i = 1, 2, \ldots, L \), we have

\[
\sigma_i \approx z_i\|W^HA_R(\theta_R,i)\|_2\|X^HA_T(\theta_T,i)\|_2,
\]

\[
u_i \approx W^HA_R(\theta_R,i)/\|W^HA_R(\theta_R,i)\|_2,
\]

\[
v_i \approx X^HA_T(\theta_T,i)/\|X^HA_T(\theta_T,i)\|_2,
\]

where \( u_i \) and \( v_i \) are the \( i \)-th column of \( U \) and \( V \), respectively, \( \{1,2,\ldots,L\} \) is a permutation of \( \{1,2,\ldots,L\} \).

By quantizing the normalized spatial angles with \( \{n\}/N_R, n = 1, \ldots, N_R \), \( \{n\}/N_T, n = 1, \ldots, N_T \), and assuming that the AoAs/AoDs just lie in the quantized points, we have the virtual representation as

\[
u_i \approx \text{D}_R a_R^{(v)}(\theta_R,i), v_i \approx X^TD_T a_T^{(v)}(\theta_T,i),
\]

where \( \text{D}_R \in \mathbb{C}^{N_R \times N_R}, \text{D}_T \in \mathbb{C}^{N_T \times N_T} \) are discrete Fourier transform (DFT) matrices, \( a_R^{(v)}, a_T^{(v)} \) have only one non-zero element. Then, we can obtain an quantized angle estimate.

We explain Algorithm 2 in more details as follows. At the beginning, we do not know the number of paths of the physical mmWave channel. Since some false paths will be cut off in the IR procedure, we need to detect some more paths than the real channel, i.e., we set \( N_{\text{init}} \) to ensure that \( L < N_{\text{init}} \) in most cases. Then, we run SVD to get \( N_{\text{init}} \) left singular vectors \( u_i \) and \( N_{\text{init}} \) right singular vectors \( v_i \) (steps 1 and 2). After that, we obtain the on-grid coarse estimate of AoAs/AoDs by finding the best matching grids (steps 4-7). After the SVD-based preconditioning as shown in Algorithm 2, the coarse estimates will be used as the initial candidates of Algorithm 1, then Algorithm 1 can be realized with much reduced computational complexity.

### Algorithm 2 SVD-based preconditioning

**Input:** Noisy received signals \( Y \), transmit pilot signals \( X \), combining matrix \( W \), and \( N_{\text{init}} \), the number of paths to detect.

**Output:** Coarse AoAs/AoDs estimates of the \( N_{\text{init}} \) paths.

1. \( [U, \Sigma, V] = \text{SVD}(Y) \).
2. Take the first \( N_{\text{init}} \) columns, \( \{u_1, u_2, \ldots, u_{N_{\text{init}}}\} \) from \( U \), and \( \{v_1, v_2, \ldots, v_{N_{\text{init}}}\} \) from \( V \), which are correspondent to the \( N_{\text{init}} \) largest singular values.
3. for \( i = 1, 2, \ldots, N_{\text{init}} \) do
   4. calculate the correlation \( c_n = (W^HD_R)^T u_i \).
   5. \( \theta_R^{(i)} = \arg\max_{j=1}^{N_R} c_n(j) - 1/N_R \).
   6. calculate the correlation \( c_n = (X^TD_T)^T v_i \).
   7. \( \theta_T^{(i)} = \arg\max_{j=1}^{N_T} c_n(j) - 1/N_T \).
8. end for

### IV. Simulation Results

In this section, simulation results are provided to investigate the performance of the proposed IR-based super-resolution channel estimation. We consider the ULA-based mmWave massive MIMO system with hybrid precoding, where \( d = \lambda/2, N_R = N_T = 64, N_{\text{RF}} = N_{\text{RF}} = 4, L = 3, \) and \( N_X = N_Y = 24 \). Then the adaptive codebook-based channel estimation [6] and the auxiliary beam pair based channel estimation [7] are adopted for performance comparison.

Fig. 1 compares the normalized mean square error (NMSE) performance against the signal-to-noise ratio (SNR). It is clear that the proposed scheme achieves much better NMSE performance. This is because the accuracy of the adaptive codebook-based channel estimation scheme [6] is limited by the codebook size, and the angle estimates are chosen in a finite set, so it suffers from an obvious error floor. The auxiliary beam pair based scheme [7] scans a channel path with a pair of beams, and obtains better angle estimation by the amplitude ratio of the received signal, but the angle estimation suffers from interference caused by other paths. By contrast, the proposed super-resolution channel estimation scheme optimizes all AoAs/AoDs simultaneously, and the interference between different paths can be eliminated, so the performance is only affected by noise.

Fig. 2 compares the average spectral efficiency of the typical hybrid precoding system [2] when different channel estimation schemes are used. The case with ideal CSI was adopted as the upper bound for performance comparison. It can be observed that the proposed super-resolution channel estimation is able to approach this upper bound. Thus, we can conclude that the proposed scheme can achieve the super-resolution channel estimation.

### V. Conclusions

In this paper, we have proposed an IR-based super-resolution channel estimation scheme for mmWave massive MIMO with hybrid precoding. Specifically, we have transformed the channel estimation problem to the optimization problem of a new objective function, which is the weighted summation of the sparsity and the data fitting error. The proposed scheme starts from the on-grid points in the angle domain, and iteratively moves them to the neighboring off-grid actual positions via gradient descent method. In addition, we
have proposed an SVD-based preconditioning to reduce the computational complexity. Simulation results have confirmed that the proposed super-resolution channel estimation scheme can accurately estimate the off-grid AoAs/AoDs with affordable complexity.

**APPENDIX A**

OPTIMIZATION OF $S$ IN (10) WITH REGARD TO $z$

For notational conciseness, we ignore the superscript (i) of $S^{(i)}$ and $D^{(i)}$ in (10), and use $A_R$, $A_T$ for $A_{R}(\theta_R)$, $A_{T}(\theta_T)$ respectively. Let $K_p = W^H A_r \text{diag} (A_R^H x_p)$. In order to find the optimal $S(z, \theta_R, \theta_T)$ with regard to $z$, we can expand the objective function $S$ as

$$S(z, \theta_R, \theta_T) = \lambda^{-1} z^H D z + \| (Y - W^H A_R \text{diag} (z) A_T^H x) \|_F^2$$

$$= \lambda^{-1} z^H D z + \sum_{p=1}^{N_x} \| y_p - W^H A_R \text{diag} (A_R^H x_p) \|_2^2$$

$$= \lambda^{-1} z^H D z + \sum_{p=1}^{N_x} \| y_p - W^H A_R \text{diag} (A_R^H x_p) \|_2^2$$

$$= \lambda^{-1} z^H D z + \sum_{p=1}^{N_x} (y_p - K_p z)^H (y_p - K_p z)$$

$$= z^H \left( \lambda^{-1} D + \sum_{p=1}^{N_x} K_p^H K_p \right) z - \sum_{p=1}^{N_x} y_p^H K_p z - \sum_{p=1}^{N_x} K_p y_p$$

Then, we can obtain the partial derivative by

$$\frac{\partial S(z, \theta_R, \theta_T)}{\partial z} = z^H \left( \lambda^{-1} D + \sum_{p=1}^{N_x} K_p^H K_p \right) - \sum_{p=1}^{N_x} y_p^H K_p$$

By setting the derivative to zero, the minimum point $z$ and the corresponding minimum value of $S(z, \theta_R, \theta_T)$ as the function of $\theta_R$ and $\theta_T$ can be obtained as

$$z_{opt}(\theta_R, \theta_T) = \left( \lambda^{-1} D + \sum_{p=1}^{N_x} K_p^H K_p \right)^{-1} \sum_{p=1}^{N_x} K_p^H y_p$$

$$S_{opt}(\theta_R, \theta_T) = - \sum_{p=1}^{N_x} K_p^H y_p^H \left( \lambda^{-1} D + \sum_{p=1}^{N_x} K_p^H K_p \right)^{-1} \sum_{p=1}^{N_x} K_p^H y_p + \sum_{p=1}^{N_x} y_p^H y_p$$

**REFERENCES**


