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

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

***Abstract*—Channel estimation is challenging for millimeter-wave massive MIMO with hybrid precoding, since the number of radio frequency 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-based superresolution 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 towards the optimal solutions, and finally realize the superresolution 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-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), superresolution 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 results in the power leakage problem, which severely degrades the channel estimation accuracy. To solve this resolution limitation caused by the on-grid angle estimation, we propose an iterative reweight (IR)-based super-resolution channel estimation scheme to estimate the off-grid AoAs/AoDs.[1]

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 the 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.

The contributions of this paper are the follows. We propose a novel IR-based super-resolution channel estimation scheme for mmWave massive MIMO with hybrid precoding. Comparing with the state-of-art schemes schemes such as those in [6]–[8], we can achieve super-resolution channel estimation, which means substantially improved estimation accuracy. Moreover, the proposed SVD based preconditioning significantly reduces the computational complexity of the IR procedure, and makes the method practical in mmWave channel estimation.

Notation: In this paper, the boldface lower and upper-case symbols denote vectors and matrices. $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ denote the transpose, the conjugate transpose, and the inverse of a matrix, respectively. $\mathrm{diag}(\mathbf{x})$ is the diagonal matrix with the vector $\mathbf{x}$ on its diagonal. The ℓ_0-norm, ℓ_2-norm, and Frobenius norm are given by $\|\cdot\|_0$, $\|\cdot\|_2$, and $\|\cdot\|_{\mathrm{F}}$, respectively.

II. System Model

We consider a hybrid-precoding mmWave massive MIMO with arbitrary array geometry. Let N_{T}, N_{R}, $N_{\mathrm{T}}^{\mathrm{RF}}$, and $N_{\mathrm{R}}^{\mathrm{RF}}$ 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., $N_{\mathrm{T}}^{\mathrm{RF}} < N_{\mathrm{T}}$, $N_{\mathrm{R}}^{\mathrm{RF}} < N_{\mathrm{R}}$ [1]–[3]. The system model can be given by

$$\mathbf{r} = \mathbf{Q}^H \mathbf{H}\mathbf{P}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{r} \in \mathbb{C}^{N_{\mathrm{R}}^{\mathrm{RF}} \times 1}$ is the received signal, $\mathbf{Q} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}^{\mathrm{RF}}}$ is the hybrid combining matrix, $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ is the channel matrix, $\mathbf{P} \in$

Manuscript received October 28, 2017; revised March 18, 2018; accepted May 27, 2018. Date of publication June 1, 2018; date of current version September 17, 2018. This work was supported in part by the National Natural Science Foundation of China for Outstanding Young Scholars under Grant 61722109, in part by the National Natural Science Foundation of China under Grant 61571270, and in part by the Royal Academy of Engineering through the U.K.–China Industry Academia Partnership Programme Scheme under Grant UK-CIAPP\49. The review of this paper was coordinated by Prof. S. Coleri Ergen. *(Corresponding author: Linglong Dai.)*

C. Hu, L. Dai, and T. Mir are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: huc16@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn; bah15@mails.tsinghua.edu.cn).

Z. Gao is with the Advanced Research Institute for Multidisciplinary Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: gaozhen16@bit.edu.cn).

J. Fang is with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: JunFang@uestc.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2842724

[1]Simulation codes are provided to reproduce the results presented in this paper: http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html

$\mathbb{C}^{N_{\rm T}\times N_{\rm T}^{\rm RF}}$ is the hybrid precoding matrix, $\mathbf{s}\in\mathbb{C}^{N_{\rm T}^{\rm RF}\times 1}$ is the transmitted signal, and $\mathbf{n}\in\mathbb{C}^{N_{\rm R}^{\rm RF}\times 1}$ is the received noise after combining.

The channel model

$$\mathbf{H}=\sum_{l=1}^{L} z_l\,\mathbf{a}_{\rm R}\left(\phi_{{\rm R},l}^{\rm azi},\phi_{{\rm R},l}^{\rm ele}\right)\mathbf{a}_{\rm T}^{H}\left(\phi_{{\rm T},l}^{\rm azi},\phi_{{\rm T},l}^{\rm ele}\right) \tag{2}$$

is widely adopted in mmWave massive MIMO systems, and it is regarded almost unchanged within the channel coherence time for channel estimation [6]–[8], where L is the number of propagation paths, $L\ll\min\left(N_{\rm R},N_{\rm T}\right)$, z_l, $\phi_{{\rm R},l}^{\rm azi}$ ($\phi_{{\rm R},l}^{\rm ele}$) and $\phi_{{\rm T},l}^{\rm azi}$ ($\phi_{{\rm T},l}^{\rm ele}$) are the complex path gain, the azimuth (elevation) AoA and AoD of the l-th path, respectively. $\mathbf{a}_{\rm R}(\phi_{{\rm R},l}^{\rm azi},\phi_{{\rm R},l}^{\rm ele})$ and $\mathbf{a}_{\rm T}(\phi_{{\rm T},l}^{\rm azi},\phi_{{\rm T},l}^{\rm ele})$ are the steering vector at the receiver and the steering vector at the transmitter, respectively. These steering vectors depend on the array geometry. Ignoring the subscripts without loss of generality, for the typical $N_1\times N_2$ uniform planar arrays (UPAs), $\mathbf{a}\left(\phi_l^{\rm azi},\phi_l^{\rm ele}\right)$ is given by [2]

$$\begin{aligned}\mathbf{a}\left(\phi^{\rm azi},\phi^{\rm ele}\right)=&\left[1,e^{j2\pi\, d\sin\phi^{\rm azi}\sin\phi^{\rm ele}/\lambda},\cdots,e^{j2\pi\,(N_1-1)d\sin\phi^{\rm azi}\sin\phi^{\rm ele}/\lambda}\right]^T\\&\otimes\left[1,e^{j2\pi\, d\cos\phi^{\rm ele}/\lambda},\cdots,e^{j2\pi\,(N_2-1)d\cos\phi^{\rm ele}/\lambda}\right]^T,\end{aligned} \tag{3}$$

where d is the antenna spacing, λ is the wavelength, $\otimes$ denotes the Kronecker product. For uniform linear arrays (ULAs), the steering vector is only determined by one angle [6]

$$\mathbf{a}\left(\phi\right)=\left[1,e^{j2\pi\, d\sin\phi/\lambda},\cdots,e^{j2\pi\,(N-1)d\sin\phi/\lambda}\right]^T. \tag{4}$$

By defining the normalized spacial angles by $\theta^{\rm azi}\triangleq d\sin\phi^{\rm azi}\sin\phi^{\rm ele}/\lambda$ and $\theta^{\rm ele}\triangleq d\cos\phi^{\rm ele}/\lambda$, the mmWave channel matrix $\mathbf{H}$ in (2) can be also written as

$$\mathbf{H}=\mathbf{A}_{\rm R}\left(\boldsymbol{\theta}_{\rm R}\right)\mathrm{diag}\left(\mathbf{z}\right)\mathbf{A}_{\rm T}^{H}\left(\boldsymbol{\theta}_{\rm T}\right), \tag{5}$$

where $\mathbf{z}=[z_1,z_2,\ldots,z_L]^T$, $\boldsymbol{\theta}_{\rm R}=[\theta_{{\rm R},1}^{\rm azi},\theta_{{\rm R},1}^{\rm ele},\theta_{{\rm R},2}^{\rm azi},\theta_{{\rm R},2}^{\rm ele},\ldots,\theta_{{\rm R},L}^{\rm azi},\theta_{{\rm R},L}^{\rm ele}]^T$, $\boldsymbol{\theta}_{\rm T}=[\theta_{{\rm T},1}^{\rm azi},\theta_{{\rm T},1}^{\rm ele},\theta_{{\rm T},2}^{\rm azi},\theta_{{\rm T},2}^{\rm ele},\ldots,\theta_{{\rm T},L}^{\rm azi},\theta_{{\rm T},L}^{\rm ele}]^T$, $\mathbf{A}_{\rm R}(\boldsymbol{\theta}_{\rm R})=\left[\mathbf{a}_{\rm R}(\theta_{{\rm R},1}^{\rm azi},\theta_{{\rm R},1}^{\rm ele})\,\mathbf{a}_{\rm R}(\theta_{{\rm R},2}^{\rm azi},\theta_{{\rm R},2}^{\rm ele})\cdots\mathbf{a}_{\rm R}(\theta_{{\rm R},L}^{\rm azi},\theta_{{\rm R},L}^{\rm ele})\right]$, $\mathbf{A}_{\rm T}(\boldsymbol{\theta}_{\rm T})=\left[\mathbf{a}_{\rm T}(\theta_{{\rm T},1}^{\rm azi},\theta_{{\rm T},1}^{\rm ele})\,\mathbf{a}_{\rm T}(\theta_{{\rm T},2}^{\rm azi},\theta_{{\rm T},2}^{\rm ele})\cdots\mathbf{a}_{\rm T}(\theta_{{\rm T},L}^{\rm azi},\theta_{{\rm T},L}^{\rm ele})\right]$.

Denote $\mathbf{x}=\mathbf{P}\mathbf{s}\in\mathbb{C}^{N_{\rm T}\times 1}$, where the i-th element of $\mathbf{x}$ is the transmitted signal at the i-th transmit antenna. Suppose that the transmitter sends $N_{\rm X}$ ($N_{\rm X}<N_{\rm T}$) different pilot sequences $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{N_{\rm X}}$. Since the number of RF chains is smaller than the required dimension of received pilot sequence, for each transmit pilot sequence $\mathbf{x}_p$ ($1\le p\le N_{\rm X}$), we use M time slots to obtain an $N_{\rm Y}$-dimensional received pilot sequence $\mathbf{y}_p$, where $N_{\rm Y}=MN_{\rm R}^{\rm RF}$. Thus, the training overhead is $T=MN_{\rm X}$. In the m-th time slot, we use the combining matrix $\mathbf{W}_m$ to obtain an $N_{\rm R}^{\rm RF}$-dimensional received pilot sequence

$$\mathbf{y}_{p,m}=\mathbf{W}_m^H\mathbf{H}\mathbf{x}_p+\mathbf{n}_{p,m}. \tag{6}$$

By collecting the received pilots in the M time slots, we have $\mathbf{y}_p=\mathbf{W}^H\mathbf{H}\mathbf{x}_p+\mathbf{n}_p$, where $\mathbf{y}_p=[\mathbf{y}_{p,1}^T,\mathbf{y}_{p,2}^T,\ldots,\mathbf{y}_{p,M}^T]^T\in\mathbb{C}^{N_{\rm Y}\times 1}$, $\mathbf{W}=[\mathbf{W}_1,\mathbf{W}_2,\ldots,\mathbf{W}_M]\in\mathbb{C}^{N_{\rm R}\times N_{\rm Y}}$, $\mathbf{n}_p\in\mathbb{C}^{N_{\rm Y}\times 1}$ is the noise. By defining $\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_{N_{\rm X}}]$, $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{N_{\rm X}}]$, $\mathbf{N}=[\mathbf{n}_1,\mathbf{n}_2,ldots,\mathbf{n}_{N_{\rm X}}]$, we have

$$\mathbf{Y}=\mathbf{W}^H\mathbf{H}\mathbf{X}+\mathbf{N}. \tag{7}$$

The estimation of the channel matrix $\mathbf{H}$ in (7) is equivalent to the estimation of the number of paths, the normalized spacial angles ($\boldsymbol{\theta}_{\rm T}$, $\boldsymbol{\theta}_{\rm R}$), and path gains $\mathbf{z}$ for all L paths. Due to the angle-domain sparsity of the channel matrix $\mathbf{H}$, the sparse channel estimation problem can be formulated as

$$\min_{\hat{\mathbf{z}},\hat{\boldsymbol{\theta}}_{\rm R},\hat{\boldsymbol{\theta}}_{\rm T}}\ \|\hat{\mathbf{z}}\|_0,\ \text{s.t.}\left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_{\rm F}\le\varepsilon, \tag{8}$$

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

III. Proposed IR-Based Super-Resolution Channel Estimation

A. Proposed Optimization Formulation

The main difficulty in solving (8) 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

$$\min_{\mathbf{z},\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T}}\ F(\mathbf{z})\triangleq\sum_{l=0}^{L}\log\left(|z_l|^2+\delta\right),\ \text{s.t.}\left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_{\rm F}\le\varepsilon, \tag{9}$$

where $\delta>0$ ensures that the logarithmic function is well-defined [10], $\hat{\mathbf{H}}$ is determined by the parameters $\mathbf{z},\boldsymbol{\theta}_{\rm R}$ and $\boldsymbol{\theta}_{\rm T}$ defined in (5). By adding a regularization parameter $\lambda>0$, we can further formulate the problem (9) as a unconstrained optimization problem:

$$\begin{aligned}\min_{\mathbf{z},\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T}}\ G\left(\mathbf{z},\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T}\right)\triangleq&\sum_{l=1}^{L}\log\left(|z_l|^2+\delta\right)\\&+\lambda\left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_{\rm F}^2.\end{aligned} \tag{10}$$

Moreover, by using an iterative surrogate function instead of the log-sum function, the minimization of $G\left(\mathbf{z},\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T}\right)$ is equivalent to the minimization of the surrogate function [10]:

$$\min_{\mathbf{z},\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T}}\ S^{(i)}(\mathbf{z},\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T})\triangleq\lambda^{-1}\mathbf{z}^{\rm H}\mathbf{D}^{(i)}\mathbf{z}+\left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_{\rm F}^2, \tag{11}$$

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

$$\mathbf{D}^{(i)}\triangleq\mathrm{diag}\left(\frac{1}{\left|\hat{z}_1^{(i)}\right|^2+\delta}\,\frac{1}{\left|\hat{z}_2^{(i)}\right|^2+\delta}\cdots\frac{1}{\left|\hat{z}_L^{(i)}\right|^2+\delta}\right), \tag{12}$$

where $\hat{\mathbf{z}}^{(i)}$ is the estimate of $\mathbf{z}$ at the i-th iteration.

Then, as proved in Appendix A, we can optimize (11) with regard to the path gains $\mathbf{z}$, to find the optimal point of $\hat{\mathbf{z}}$ and the corresponding optimal value of $S^{(i)}$ as follows:

$$\begin{aligned}\mathbf{z}_{\rm opt}^{(i)}(\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T})\triangleq&\ \arg\min_{\mathbf{z}}\ S^{(i)}(\mathbf{z},\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T})\\=&\left(\lambda^{-1}\mathbf{D}^{(i)}+\sum_{p=1}^{N_{\rm X}}\mathbf{K}_p^H\mathbf{K}_p\right)^{-1}\left(\sum_{p=1}^{N_{\rm X}}\mathbf{K}_p^H\mathbf{y}_p\right),\end{aligned} \tag{13}$$

$$\begin{aligned}S_{\rm opt}^{(i)}(\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T})\triangleq&\ \min_{\mathbf{z}}\ S^{(i)}(\mathbf{z},\boldsymbol{\theta}_{\rm R},\boldsymbol{\theta}_{\rm T})\\=&-\left(\sum_{p=1}^{N_{\rm X}}\mathbf{y}_p^H\mathbf{K}_p\right)\cdot\left(\lambda^{-1}\mathbf{D}^{(i)}+\sum_{p=1}^{N_{\rm X}}\mathbf{K}_p^H\mathbf{K}_p\right)^{-1}\\&\cdot\left(\sum_{p=1}^{N_{\rm X}}\mathbf{K}_p^H\mathbf{y}_p\right)+\sum_{p=1}^{N_{\rm X}}\mathbf{y}_p^H\mathbf{y}_p,\end{aligned} \tag{14}$$

where $\mathbf{K}_p=\mathbf{W}^H\mathbf{A}_{\rm R}\,\mathrm{diag}\left(\mathbf{A}_{\rm T}^H\mathbf{x}_p\right)$. After that, we only need to optimize the normalized spatial angles $\boldsymbol{\theta}_{\rm R}$ and $\boldsymbol{\theta}_{\rm T}$ in (14), which will be discussed in the next section.

Algorithm 1: IR-based Super-Resolution Channel Estimation.

Input: Noisy received signals $\mathbf{Y}$, transmit pilot signals $\mathbf{X}$, combining matrix $\mathbf{W}$, initial on-grid AoAs and AoDs $\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(0)}$, $\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(0)}$, pruning threshold z_{th} and termination threshold $\varepsilon_{\mathrm{th}}$.

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

1: Initialize $\hat{\mathbf{z}}^{(0)} = \mathbf{z}_{\mathrm{opt}}\left(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(0)}, \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(0)}\right)$ according to (13).
2: **repeat**
3: Update λ by (15).
4: Construct the function $S_{\mathrm{opt}}^{(i)}(\boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}})$ by (14).
5: Search for new angle estimates $\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i+1)}$, $\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i+1)}$ by (17).
6: Estimate the path gains $\hat{\mathbf{z}}^{(i+1)}$ according to (13).
7: Prune path l if $\hat{z}_l^{(i+1)} < z_{\mathrm{th}}$.
8: **until** $L^{(i)} = L^{(i+1)}$ and $\left\|\mathbf{z}^{(i+1)} - \mathbf{z}^{(i)}\right\|_2 < \varepsilon_{\mathrm{th}}$.
9: $\hat{\boldsymbol{\theta}}_{\mathrm{R}} = \hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(\mathrm{last})}, \hat{\boldsymbol{\theta}}_{\mathrm{T}} = \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(\mathrm{last})}, \hat{\mathbf{z}} = \hat{\mathbf{z}}^{(\mathrm{last})}$.

Algorithm 2: SVD-based preconditioning.

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

Output: Coarse AoAs/AoDs estimates of the N_{init} paths.

1: $[\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}] = \mathrm{SVD}(\mathbf{Y})$.
2: Take the first N_{init} columns, $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{N_{\mathrm{init}}}\}$ from $\mathbf{U}$, and $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{N_{\mathrm{init}}}\}$ from $\mathbf{V}$, which are correspondent to the N_{init} largest singular values.
3: **for** $i = 1, 2, \cdots, N_{\mathrm{init}}$ **do**
4: $\left(\hat{\theta}_{\mathrm{R},i}^{\mathrm{azi}(0)}, \hat{\theta}_{\mathrm{R},i}^{\mathrm{ele}(0)}\right) = \underset{\left(\theta_{\mathrm{R}}^{\mathrm{azi}}, \theta_{\mathrm{R}}^{\mathrm{ele}}\right) \in \Omega_{\mathrm{R}}}{\arg\max} \mathbf{u}_i^H \mathbf{W}^H \mathbf{a}_{\mathrm{R}}\left(\theta_{\mathrm{R}}^{\mathrm{azi}}, \theta_{\mathrm{R}}^{\mathrm{ele}}\right)$.
5: $\left(\hat{\theta}_{\mathrm{T},i}^{\mathrm{azi}(0)}, \hat{\theta}_{\mathrm{T},i}^{\mathrm{ele}(0)}\right) = \underset{\left(\theta_{\mathrm{T}}^{\mathrm{azi}}, \theta_{\mathrm{T}}^{\mathrm{ele}}\right) \in \Omega_{\mathrm{T}}}{\arg\max} \mathbf{v}_i^H \mathbf{X}^H \mathbf{a}_{\mathrm{T}}\left(\theta_{\mathrm{T}}^{\mathrm{azi}}, \theta_{\mathrm{T}}^{\mathrm{ele}}\right)$.
6: **end for**

B. IR-Based Super-Resolution Channel Estimation

In the previous section, we have already simplified the constrained optimization problem (8) to an unconstrained angle optimization problem (14). To solve this reformulated problem, now we propose an IR-based super-resolution channel estimation scheme as described in Algorithm 1.

The objective function $S^{(i)}(\mathbf{z}, \boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}})$ is the weighted sum of two parts: $\mathbf{z}^{\mathrm{H}}\mathbf{D}\mathbf{z}$ controlling the sparsity of the estimation result and $\|\mathbf{Y} - \mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\|_{\mathrm{F}}$ denoting the residue. In addition, λ is the regularization parameter that controls the tradeoff between the sparsity and the data fitting error.

In the iterative reweighted method [10], λ is not fixed but updated in each iteration. To be specific, if the previous iteration is poorly-fitted, we will choose a smaller λ 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 λ to accelerate the searching for the best-fitting estimate. In the proposed algorithm, λ is updated by

$$\lambda = \min(d/r^{(i)}, \lambda_{\max}), \tag{15}$$

where d is a constant scaling factor, and $\lambda_{\max}$ is selected to make the problem well-conditioned, $r^{(i)}$ is the squared residue in the previous step, i.e.,

$$r^{(i)} = \|\mathbf{Y} - \mathbf{W}^H\mathbf{A}_{\mathrm{R}}(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i)})\,\mathrm{diag}(\hat{\mathbf{z}}^{(i)})\mathbf{A}_{\mathrm{T}}^H(\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i)})\mathbf{X}\|_{\mathrm{F}}^2. \tag{16}$$

The updation of λ was discussed in [10] with more details.

The proposed algorithm starts iteration at the angle domain grids. In the i-th iteration, our task is to search for new estimates $\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i+1)}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i+1)}$ in the neighborhood of the previous estimates $\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i)}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i)}$ to make the objective function $S^{(i)}$ become smaller. This searching can be accomplished via gradient descent method:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i+1)} &= \hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i)} - \eta \cdot \nabla_{\boldsymbol{\theta}_{\mathrm{R}}} S_{\mathrm{opt}}^{(i)}\left(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i)}, \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i)}\right),\\ \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i+1)} &= \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i)} - \eta \cdot \nabla_{\boldsymbol{\theta}_{\mathrm{T}}} S_{\mathrm{opt}}^{(i)}\left(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i)}, \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i)}\right),\end{aligned} \tag{17}$$

where the gradients can be calculated according to Appendix B, and η is the chosen step-length to make sure $S_{\mathrm{opt}}^{(i)}(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i+1)}, \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i+1)}) \leq S_{\mathrm{opt}}^{(i)}(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(i)}, \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(i)})$. The estimates become more and more accurate during the iterative searching, until the new estimates are almost the same as the previous ones. With our proposed IR-based super-resolution channel estimation scheme, the estimates of $(\boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}})$ 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.

The computational complexity in each iteration lies in calculating the gradient in Step 5. The computational complexity to calculate the gradient is $\mathcal{O}\left(N_{\mathrm{X}}N_{\mathrm{Y}}(N_{\mathrm{R}}+N_{\mathrm{T}})L^2\right)$. 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 $\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(0)}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(0)}$ 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. Comparing to using all $N_{\mathrm{R}}N_{\mathrm{T}}$ angle-domain grids as initial candidates, the preconditioning can significantly reduce the computational complexity of the IR-based super-resolution channel estimation scheme.

Specifically, by applying SVD to the matrix $\mathbf{Y}$, we have $\mathbf{Y} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, where $\boldsymbol{\Sigma} = \mathrm{diag}\left(\sigma_1, \sigma_2, \ldots, \sigma_{\min(N_{\mathrm{X}}, N_{\mathrm{Y}})}\right) \in \mathbb{R}^{N_{\mathrm{Y}} \times N_{\mathrm{X}}}$ whose diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(N_{\mathrm{X}}, N_{\mathrm{Y}})} \geq 0$ are the singular values of $\mathbf{Y}$, and $\mathbf{U}^H\mathbf{U} = \mathbf{I}_{N_{\mathrm{Y}} \times N_{\mathrm{Y}}}$, $\mathbf{V}^H\mathbf{V} = \mathbf{I}_{N_{\mathrm{X}} \times N_{\mathrm{X}}}$. From (5) and (7), we have

$$\mathbf{Y} = \left(\mathbf{W}^H\mathbf{A}_{\mathrm{R}}(\boldsymbol{\theta}_{\mathrm{R}})\right)\mathrm{diag}(\mathbf{z})\left(\mathbf{X}^H\mathbf{A}_{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{T}})\right)^H + \mathbf{N}. \tag{18}$$

As the noise is small, the largest L singular values and their corresponding singular vectors are approximately determined by the L

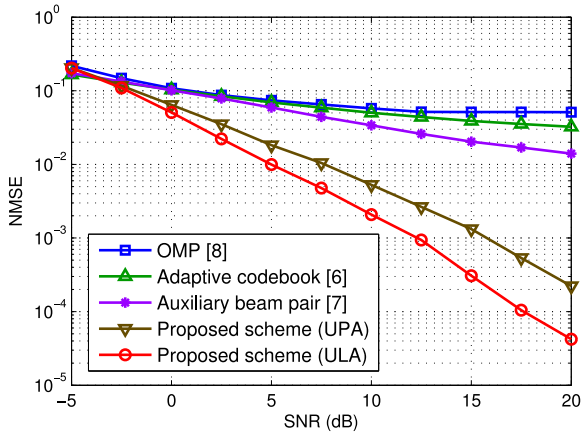

Fig. 1. NMSE performance comparison of different channel estimation schemes under NLoS channel.

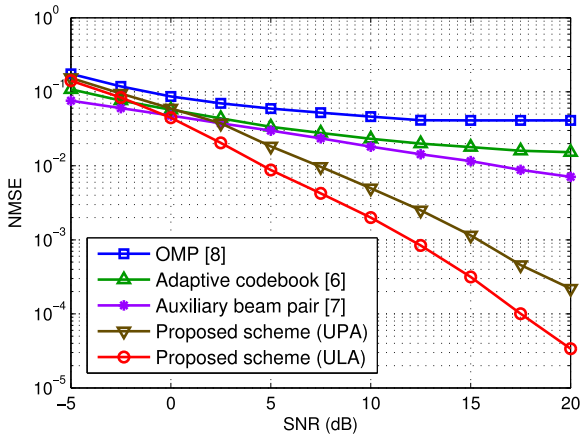

Fig. 2. NMSE performance comparison of different channel estimation schemes under LoS channel.

paths, i.e., for $i = 1, 2, \cdots, L$, we have

$$\begin{aligned}
\sigma_i &\approx |z_{l_i}| \left\|\mathbf{W}^H \mathbf{a}_{\mathrm{R}}\left(\theta_{\mathrm{R},l_i}^{\mathrm{azi}}, \theta_{\mathrm{R},l_i}^{\mathrm{ele}}\right)\right\|_2 \left\|\mathbf{X}^H \mathbf{a}_{\mathrm{T}}\left(\theta_{\mathrm{T},l_i}^{\mathrm{azi}}, \theta_{\mathrm{T},l_i}^{\mathrm{ele}}\right)\right\|_2, \\
\mathbf{u}_i &\approx \mathbf{W}^H \mathbf{a}_{\mathrm{R}}\left(\theta_{\mathrm{R},l_i}^{\mathrm{azi}}, \theta_{\mathrm{R},l_i}^{\mathrm{ele}}\right) / \left\|\mathbf{W}^H \mathbf{a}_{\mathrm{R}}\left(\theta_{\mathrm{T},l_i}^{\mathrm{azi}}, \theta_{\mathrm{T},l_i}^{\mathrm{ele}}\right)\right\|_2, \\
\mathbf{v}_i &\approx \mathbf{X}^H \mathbf{a}_{\mathrm{T}}\left(\theta_{\mathrm{R},l_i}^{\mathrm{azi}}, \theta_{\mathrm{R},l_i}^{\mathrm{ele}}\right) / \left\|\mathbf{X}^H \mathbf{a}_{\mathrm{T}}\left(\theta_{\mathrm{T},l_i}^{\mathrm{azi}}, \theta_{\mathrm{T},l_i}^{\mathrm{ele}}\right)\right\|_2,
\end{aligned} \tag{19}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th column of $\mathbf{U}$ and $\mathbf{V}$, respectively, $\{l_1, l_2, \cdots, l_L\}$ is a permutation of $\{1, 2, \cdots, L\}$.

Then, in steps 4-5 of Algorithm 2, we search for the coarse estimate of normalized AoAs (AoDs) in the finite set of angle-domain grids Ω_{R} (Ω_{T}). Take UPA as an example. For an $N_1 \times N_2$ receiver array, the set of grids can be defined by $\Omega_{\mathrm{R}} = \{(i/N_1, j/N_2) | i = 0, 1, \ldots, N_1 - 1; j = 0, 1, \ldots, N_2 - 1\}$. We can similarly define Ω_{T} for the transmitter.

In [10], the initial candidates of Algorithm 1 are set to be all the grids, i.e., $L^{(0)} = N_{\mathrm{R}} N_{\mathrm{T}}$. The computational complexity is $\mathcal{O}\left(N_{\mathrm{X}} N_{\mathrm{Y}} (N_{\mathrm{R}} + N_{\mathrm{T}}) N_{\mathrm{R}}^2 N_{\mathrm{T}}^2\right)$, which is unaffordable when N_{R} and N_{T} are very large. Fortunately, with the proposed SVD-based preconditioning shown in Algorithm 2, the coarse estimates will be used as the initial candidates of Algorithm 1, i.e., $L^{(0)} = N_{\mathrm{init}} \approx L$. Thus, the computational complexity after SVD preconditioning is $\mathcal{O}\left(N_{\mathrm{X}} N_{\mathrm{Y}} (N_{\mathrm{R}} + N_{\mathrm{T}}) L^2\right)$, which is much lower than directly applying the scheme in [10].

IV. Simulation Results and Discussions

In this section, simulation results are provided to investigate the performance. We consider the mmWave massive MIMO system with hybrid precoding, where $L = 3$, $d = \lambda/2$, $N_{\mathrm{R}} = N_{\mathrm{T}} = 64$, $N_{\mathrm{R}}^{\mathrm{RF}} = N_{\mathrm{T}}^{\mathrm{RF}} = 4$ and $N_{\mathrm{X}} = N_{\mathrm{Y}} = 32$. The path gains are assumed Gaussian, i.e., $\alpha_l \sim \mathcal{CN}(0, \sigma_\alpha^2)$. Each element of the transmitted pilots $\mathbf{X}$ satisfies $x_{i,j} = \sqrt{\rho / N_{\mathrm{T}}} e^{j\omega_{i,j}}$, where ρ is the transmitted power, $\omega_{i,j}$ is the random phase uniformly distributed in $[0, 2\pi)$. The signal-to-noise ratio (SNR) is defined by $\mathrm{SNR} = \frac{\rho \sigma_\alpha^2}{\sigma_n^2}$, where σ_n^2 is the noise variance. We consider the ULA geometry, so that the adaptive codebook-based channel estimation [6], the auxiliary beam pair based channel estimation [7], and the OMP-based channel estimation [8] can be adopted for performance comparison.

Figs. 1 and 2 compares the normalized mean square error (NMSE) performance against SNR, under none-line-of-sight (NLoS) and line-of-sight (LoS) channels, respectively. The Rician K-factor is 20 dB in the LoS scenario. In both cases, the proposed scheme achieves much better NMSE performance when SNR becomes large. Moreover, we show the performance of the proposed scheme when UPA is considered. Both the transmitter and the receiver adopt 64-antenna UPA with 8 rows and 8 columns. We estimate the azimuth and elevation angles at both sides. We can observe that the proposed scheme is also able to achieve super-resolution channel estimation when UPA is used. Since the estimation errors of both azimuth and elevation angles contribute to the NMSE, under the same number of antennas and number of pilot overhead, the NMSE performance of UPA is higher than that of ULA.

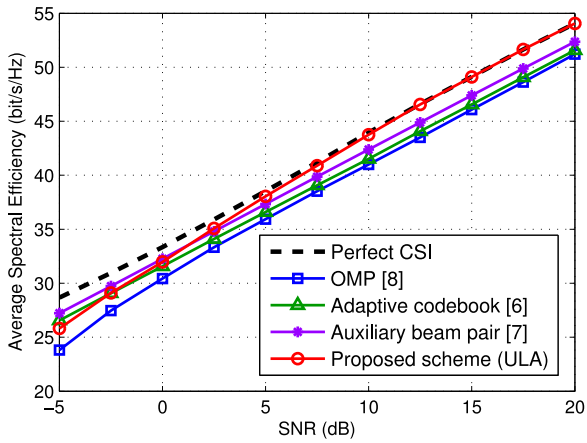

Fig. 3. Average spectral efficiency when different channel estimation schemes are used.

Fig. 3 compares the average spectral efficiency when different channel estimation schemes are used. The spectral efficiency is evaluated in the hybrid precoding system [2]. 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. This is because the angle resolution of the proposed scheme does not suffer from limited codebook size or angle quantization. Thus, we can conclude that the proposed scheme can achieve the super-resolution channel estimation.

There is a tradeoff between the channel estimation accuracy and the computation complexity. The proposed scheme with SVD preconditioning is able to achieve much higher channel estimation accuracy, but has a higher computational complexity. The computational complexity

of the proposed super-resolution channel estimation scheme with SVD preconditioning is $\mathcal{O}\left(N_{\mathrm{X}} N_{\mathrm{Y}} (N_{\mathrm{R}}+N_{\mathrm{T}}) L^2\right)$. In comparison, the computational complexity of the OMP-based channel estimation [8] is $\mathcal{O}\left(N_{\mathrm{X}} N_{\mathrm{Y}} (N_{\mathrm{R}}+N_{\mathrm{T}}) L\right)$. In order to achieve higher estimation accuracy, the increase in computational complexity is acceptable since L is usually small for mmWave channels.

V. Conclusion

In this correspondence, 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 advance the state-of-art by estimating the off-grid AoAs/AoDs with much increased accuracy. Angle estimation is the key of channel estimation for mmWave massive MIMO. Estimating the AoAs/AoDs with higher resolution is a practical way to realize higher spectral efficiency. For future work, it would be interesting to study other super-resolution channel estimation schemes with reduced complexity. In addition, super-resolution channel estimation under high mobility is an important yet challenging topic to be investigated.

Appendix A
Optimization of S in (11) With Regard to $\mathbf{z}$

For notational conciseness, we ignore the superscript (i) of $S^{(i)}$ and $\mathbf{D}^{(i)}$ in (11), and use $\mathbf{A}_{\mathrm{R}}$, $\mathbf{A}_{\mathrm{T}}$ for $\mathbf{A}_{\mathrm{R}}(\boldsymbol{\theta}_{\mathrm{R}})$, $\mathbf{A}_{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{T}})$ respectively. Let $\mathbf{K}_p = \mathbf{W}^H \mathbf{A}_r \operatorname{diag}\left(\mathbf{A}_t^{\mathrm{H}} \mathbf{x}_p\right)$. In order to find the optimal $S(\mathbf{z}, \boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}})$ with regard to $\mathbf{z}$, we can expand the objective function S as

$$\begin{aligned} S(\mathbf{z}, \boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}}) &= \lambda^{-1}\mathbf{z}^H \mathbf{D}\mathbf{z} \\ &\quad + \sum_{p=1}^{N_{\mathrm{X}}} \left\| \mathbf{y}_p - \mathbf{W}^H \mathbf{A}_{\mathrm{R}} \operatorname{diag}(\mathbf{z}) \mathbf{A}_{\mathrm{T}}^H \mathbf{x}_p \right\|_2^2 \\ &= \lambda^{-1}\mathbf{z}^H \mathbf{D}\mathbf{z} + \sum_{p=1}^{N_{\mathrm{X}}} (\mathbf{y}_p - \mathbf{K}_p \mathbf{z})^H (\mathbf{y}_p - \mathbf{K}_p \mathbf{z}) \\ &= \mathbf{z}^H \left(\lambda^{-1}\mathbf{D} + \sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{K}_p\right)\mathbf{z} - \mathbf{z}^H \left(\sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{y}_p\right) - \left(\sum_{p=1}^{N_{\mathrm{X}}} \mathbf{y}_p^H \mathbf{K}_p\right)\mathbf{z} \\ &\quad + \sum_{p=1}^{N_{\mathrm{X}}} \mathbf{y}_p^H \mathbf{y}_p, \end{aligned} \tag{20}$$

Then, we can obtain the partial derivative by

$$\frac{\partial S(\mathbf{z}, \boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}})}{\partial \mathbf{z}} = \mathbf{z}^H \left(\lambda^{-1}\mathbf{D} + \sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{K}_p\right) - \left(\sum_{p=1}^{N_{\mathrm{X}}} \mathbf{y}_p^H \mathbf{K}_p\right). \tag{21}$$

By setting the derivative to zero, the minimum point $\mathbf{z}$ and the corresponding minimum value of $S(\mathbf{z}, \boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}})$ as the function of $\boldsymbol{\theta}_{\mathrm{R}}$ and $\boldsymbol{\theta}_{\mathrm{T}}$ can be obtained as

$$\mathbf{z}_{\mathrm{opt}}(\boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}}) = \left(\lambda^{-1}\mathbf{D} + \sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{K}_p\right)^{-1} \left(\sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{y}_p\right), \tag{22}$$

$$\begin{aligned} S_{\mathrm{opt}}(\boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}}) = &- \left(\sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{y}_p\right)^H \left(\lambda^{-1}\mathbf{D} + \sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{K}_p\right)^{-1} \\ &\cdot \left(\sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{y}_p\right) + \sum_{p=1}^{N_{\mathrm{X}}} \mathbf{y}_p^H \mathbf{y}_p. \end{aligned} \tag{23}$$

Appendix B
Gradient of $S_{\mathrm{opt}}(\boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}})$

Denote $\mathbf{v} = \sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{y}_p$, $\mathbf{A} = \lambda^{-1}\mathbf{D} + \sum_{p=1}^{N_{\mathrm{X}}} \mathbf{K}_p^H \mathbf{K}_p$, we have $S_{\mathrm{opt}} = -\mathbf{v}^H \mathbf{A}^{-1}\mathbf{v} + \sum_{p=1}^{N_{\mathrm{X}}} \mathbf{y}_p^H \mathbf{y}_p$. Take partial derivative with respect to $\theta_{\mathrm{R},l}$, we have

$$\begin{aligned} \frac{\partial S_{\mathrm{opt}}}{\partial \theta_{\mathrm{R},l}} &= -\frac{\partial \mathbf{v}^H}{\partial \theta_{\mathrm{R},l}} \mathbf{A}^{-1}\mathbf{v} - \mathbf{v}^H \frac{\partial \mathbf{A}^{-1}}{\partial \theta_{\mathrm{R},l}} \mathbf{v} - \mathbf{v}^H \mathbf{A}^{-1} \frac{\partial \mathbf{v}}{\partial \theta_{\mathrm{R},l}} \\ &= -\frac{\partial \mathbf{v}^H}{\partial \theta_{\mathrm{R},l}} \mathbf{A}^{-1}\mathbf{v} + \mathbf{v}^H \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \theta_{\mathrm{R},l}} \mathbf{A}^{-1}\mathbf{v} - \mathbf{v}^H \mathbf{A}^{-1} \frac{\partial \mathbf{v}}{\partial \theta_{\mathrm{R},l}}, \end{aligned} \tag{24}$$

where

$$\frac{\partial \mathbf{v}}{\partial \theta_{\mathrm{R},l}} = \sum_{p=1}^{N_{\mathrm{X}}} \frac{\partial \mathbf{K}_p^H}{\partial \theta_{\mathrm{R},l}} \mathbf{y}_p, \quad \frac{\partial \mathbf{A}}{\partial \theta_{\mathrm{R},l}} = \sum_{p=1}^{N_{\mathrm{X}}} \left(\frac{\partial \mathbf{K}_p^H}{\partial \theta_{\mathrm{R},l}} \mathbf{K}_p + \mathbf{K}_p^H \frac{\partial \mathbf{K}_p}{\partial \theta_{\mathrm{R},l}}\right), \tag{25}$$

$$\frac{\partial \mathbf{K}_p}{\partial \theta_{\mathrm{R},l}} = \left[\mathbf{0} \cdots \mathbf{0}\ \mathbf{W}^H \frac{\partial \mathbf{a}_{\mathrm{R}}(\theta_{\mathrm{R},l})}{\partial \theta_{\mathrm{R},l}} \mathbf{a}_{\mathrm{T}}^H(\theta_{\mathrm{T},l})\, \mathbf{x}_p\ \mathbf{0} \cdots \mathbf{0}\right].$$

References

[1] S. Mumtaz, J. Rodriquez, and L. Dai, *MmWave Massive MIMO: A Paradigm for 5G*. New York, NY, USA: Academic, 2016.

[2] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[3] X. Gao, L. Dai, S. Han, C.-L. I, and R. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[4] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[5] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[6] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[7] D. Zhu, J. Choi, and R. W. Heath, "Auxiliary beam pair enabled AoD and AoA estimation in closed-loop large-scale millimeter-wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4770–4785, May 2017.

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

[9] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.

[10] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649–4662, Sep. 2016.