Article:

Zhang, Aihua, Cao, Wenzhou, Liu, Pengcheng orcid.org/0000-0003-0677-4421 et al. (2 more authors) (Accepted: 2020) Channel Estimation for MmWave Massive MIMO with Hybrid Precoding Based on Log-Sum Sparse Constraints. IEEE Transactions on Circuits and Systems II: Express Briefs. ISSN 1549-7747 (In Press)

Channel Estimation for MmWave Massive MIMO with Hybrid Precoding Base on Log-Sum Sparse Constraints

Aihua Zhang, *Member, IEEE,* Wenzhou Cao, Pengcheng liu, *Member, IEEE,* Jun Sun, and Jianjun Li

***Abstract*—Channel estimation is essential for millimeter-wave (mmWave) multiple-input multiple-output (MIMO) systems with hybrid precoding. However, accurate channel estimation is a challenging task as the number of antennas is huge, while the number of RF chains is limited. Traditional methods of compressed sensing for channel estimation lead to serious loss of accuracy due to channel angle quantization. In this paper, we propose a new iterative reweight-based log-sum constraint channel estimation scheme. Specifically, we exploit the structure sparsity of the mmWave channels by formulating the channel estimation problem as an objective optimization problem. We utilize the log-sum as a constraint, via optimizing an objective function through the gradient descent method, the proposed algorithm can iteratively move the channel estimated angle-of-arrivals (AOAs) and angle-of-departures (AODs) towards the optimal solutions, and finally improve the angle estimation performance significantly. In addition, to ensure the accuracy of channel estimation, we introduce a dynamic regularization factor to control the tradeoff between the channel sparsity and the data fitting error. Numerical experiments demonstrate that the proposed algorithm achieves better convergence behavior than conventional sparse signal recovery solutions.**

***Index Terms*—Channel Estimation, mmWave, MIMO, Hybrid Precoding, Log-Sum, AOAs, AODs.**

I. Introduction

MILLIMETER-wave (mmWave) massive MIMO has been considered as a dominant technology for wireless communications in recent years [1]. By utilizing hundreds or even thousands of antennas at the base station (BS), the mmWave massive MIMO systems can provide very wide spectrum bandwidths and achieve large capacity and high throughput. In practical mmWave massive MIMO systems, a large number of antenna arrays are driven by a corresponding number of radio-frequency (RF) chains that consist of digital-to-analog/analog-to-digital converters, mixers etc. This will bring unaffordable hardware cost and energy consumption. To reduce the power consumption and hardware cost, hybrid precoding architecture with much smaller number of RF chains has been considered [2]. However, the digital baseband cannot directly access all antennas with limited RF chains [3]. Therefore, it is difficult to obtain the high-dimensional channel state information accurately.

A. Zhang,W.Cao and J.Li are with Zhongyuan University of Technology, School of electronic information, Zhengzhou, China. (zhah@zut.edu.cn; wzhou0911@gmail.com; 6522@zut.edu.cn).

P. Liu is with the Department of Computer Science, University of York, York YO10 5GH, United Kingdom (pengcheng.liu@ieee.org).

J.Sun is with the School of Information Engineering, Zhengzhou University, Zhengzhou, Henan, China (sunjun@zut.edu.cn).

Recently, several channel estimation algorithms were proposed for mmWave massive MIMO systems with hybrid precoding [4]-[9]. X. Li et al. proposed a two-stage CS scheme, which utilize the sparse and low-rank properties of the angular spreads domain [4].M. J. Azizipour et al. proposed a new greedy-based algorithm for sparse channel estimation in [5], and the authors assumed that the massive channel is totally unknown and then exploited the inherent property of the correlation between the measurements and the sensing matrix to estimate the channel. To recover accurate channel state information for wideband mmWave MIMO system, a novel framework jointly exploiting the channel's low-rank and the angular information was proposed by E. Vlachos et al. in [6]. Due to the beam squint of the mmWave MIMO system, M. Wang et al. utilized the shift-invariant block sparsity of the resulting nonstandard channel model to design a compressive sensing-based channel estimation algorithm [7]. The work in [8] exploited the block sparsity of massive MIMO channels, and calculated the channel autocorrelation matrix by investigating the channel prior information based on compressive sensing (CS) theory. Then, the L.Ge et al.of [8] used regularized method to treat the channel estimation as a convex optimization problem. S. Rao et al. established performance bounds on the channel estimation of one-bit mmWave massive MIMO receivers for different types of channel models, and considered the structured of the channel model [9]. However, these kinds of estimation algorithms that employed the AoAs/AoDs of the paths, usually assumed the angle of the AoAs/AoDs is discrete. They are continuously distributed in practice, and the assumptions of the AoAs/AoDs creat power leakage problem which will lead to a certain loss of channel estimation performance.

To solve this accurate limitation caused by discrete angle distributed, in this paper, we propose an iterative reweight based on the log-sum constrained algorithm. Specifically, we take advantage of the limited scattering of the millimeter wave propagation paths, and formulate the channel estimation problem as a sparse signal optimization problem. Then we move the estimates of the AoAs/AoDs towards the optimal solution through gradient descent method. In addition, each iteration generates a suitable weight value to control the trade-off between the sparsity and the data fitting error. To reduce the computational complexity, we utilize the singular value decomposition (SVD)-based preconditioning method. At last, the novel scheme shows superior performance in comparison with the state-of-art methods through the simulation results.

The main contributions of this paper are summarized as follows. We propose a new log-sum constrained channel estimation algorithm for mmWave massive MIMO with hybrid precoding. Comparing with the IR-based scheme [12], we can achieve better estimation performance. Moreover, we introduce a dynamic regularization factor to control the tradeoff between the channel sparsity and the data fitting error, which can guarantee the potential estimation accuracy of the proposed algorithm.

The rest of the paper is organized as follows. In Section II, the system model of mmWave massive MIMO with hybrid precoding is introduced. In section III, the theoretical details of the proposed scheme are analyzed. In section IV, several experiments provided to prove the excellent performance of the proposed algorithm. Section V is the conclusion.

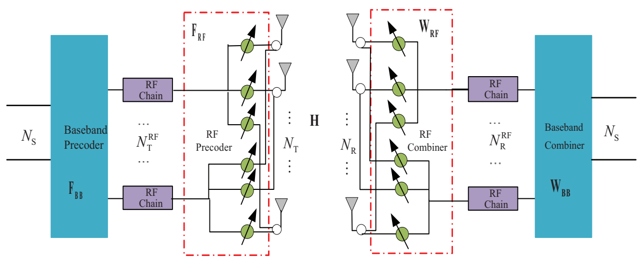


Fig. 1. Hybrid-precoding mmWave massive MIMO system model.

II. System Model

In this section, we consider a hybrid-precoding mmWave massive MIMO system model that was proposed in [10]. As shown in Fig.1, $N_{\rm T}$, $N_{\rm R}$, $N_{\rm T}^{\rm RF}$ and $N_{\rm R}^{\rm RF}$ indicate the number of the base station (BS) antennas and the mobile station (MS) antennas, the transmitter RF chains, and the receiver RF chains respectively. The BS communicates with MS through $N_{\rm s}$ data streams,such that $N_{\rm T}^{\rm RF} < N_{\rm T}$, $N_{\rm R}^{\rm RF} < N_{\rm R}$. The size of the baseband precoder $\mathbf{F}_{\rm BB}$ is $N_{\rm T}^{\rm RF} \times N_{\rm s}$, and the corresponding RF precoder $\mathbf{F}_{\rm RF}$ is $N_{\rm T} \times N_{\rm T}^{\rm RF}$. The received signal by MS can be expressed as

$$\mathbf{R} = \mathbf{H}\mathbf{F}_{\rm T}\mathbf{s} + \mathbf{n} \ , \tag{1}$$

where $\mathbf{R}$ is the $N_{\rm R} \times 1$ received signal, $\mathbf{H}$ is the $N_{\rm R} \times N_{\rm T}$ channel matrix, $\mathbf{F}_{\rm T}$=$\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}$, $\mathbf{s}$ is the $N_{\rm s} \times 1$ transmitted symbol vector, $\mathbf{n}$ is the additive white Gaussian noise vector.

In (1), supposing $\mathbf{X}$=$\mathbf{F}_{\rm T}\mathbf{s}$ is the transmitted signal. We assume that the transmitter sends $C_{\rm X}$ ($C_{\rm X} < N_{\rm T}$) pilot sequences $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{C_{\rm X}}$. Since the number of chains is much smaller than the dimension of the received pilot signals, we divide the $C_{\rm Y}$-dimensional received signal into T time intervals ($C_{\rm Y}$=$TN_{\rm R}^{\rm RF}$). At t_0 ($0 < t_0 \leq T$) time intervals, for each transmission sequence $\mathbf{x}_m$ ($1 \leq m \leq C_{\rm X}$), the corresponding reception sequence $\mathbf{y}_{m,t_0}$ can be expressed as

$$\mathbf{y}_{m,t_0} = \mathbf{W}_m^H \mathbf{H}\mathbf{x}_m + \mathbf{n}_{m,t_0} \ , \tag{2}$$

where $\mathbf{W}_m$ is the combining matrix that is consisted of the RF combining matrix $\mathbf{W}_{\rm RF}$ and the baseband combiners $\mathbf{W}_{\rm BB}$. $\mathbf{n}_{m,t_0}$ represents the noise signal at time t_0. After T time intervals, $\mathbf{y}_m = \mathbf{W}_m^H \mathbf{H}\mathbf{x}_m + \mathbf{n}_m$. $\mathbf{n}_m$ still represents the noise at the m transmission sequence. Defining $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{C_{\rm X}}]$, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_{C_{\rm X}}]$, $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_{\rm T}]$, $\mathbf{N}_o = [\mathbf{n}_1, \mathbf{n}_2, \cdots, \mathbf{n}_{C_{\rm X}}]$. we obtain

$$\mathbf{Y} = \mathbf{W}^H \mathbf{H}\mathbf{X} + \mathbf{N}_o \ . \tag{3}$$

The channel model can be expressed as

$$\mathbf{H} = \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\rm R}(\theta_l^r, \phi_l^r)\mathbf{a}_{\rm T}^H(\theta_l^t, \phi_l^r) \ . \tag{4}$$

The mmWave channel model as (4) is expected to have a geometric structure with limited scatters [11]. Where L denotes the number of propagation paths. α_l is the complex gain, $\theta_l^r(\phi_l^r)$ and $\theta_l^t(\phi_l^t)$ represent for the received and transmitted azimuth (elevation) angles of departure or arrival of the lth path, respectively. $\mathbf{a}_{\rm R}(\theta_l)$ and $\mathbf{a}_{\rm T}(\phi_l)$ are the steering vector at the MS and BS respectively, which depend on the array geometry. They have two common array structures uniform linear arrays (ULA) and uniform planar arrays (UPA) [12].

For an N-element ULA steering vector, the elevation domain is invariant, so the response vector of the ULA can be expressed as

$$\mathbf{a}_{\rm ULA}\left(\theta\right) = [1, e^{j2\pi d \sin\theta/\lambda}, \cdots, e^{j2\pi(N-1)d\sin\theta/\lambda}]^T \ . \tag{5}$$

The UPA steering vector with $N_1 \times N_2$ can be written as

$$\begin{aligned}\mathbf{a}_{\rm UPA}\left(\theta, \phi\right) = & [1, e^{j2\pi d \sin\theta\sin\phi/\lambda}, \ \cdots \\ & , e^{j2\pi(N_1-1)d\sin\theta\sin\phi/\lambda}]^T \\ & \otimes [1, e^{j2\pi d\cos\phi/\lambda}, \cdots, e^{j2\pi(N_2-1)d\cos\phi/\lambda}]^T,\end{aligned} \tag{6}$$

where $\otimes$ denotes the Kronecker product. d is the inter-element spacing, and λ is the wavelength.

The channel $\mathbf{H}$ in (3) can be also written as

$$\mathbf{H} = \mathbf{A}_{\rm R}(\boldsymbol{\theta}_{\rm R}) diag(\boldsymbol{\alpha}) \mathbf{A}_{\rm T}^H(\boldsymbol{\theta}_{\rm T}) \ , \tag{7}$$

where $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_L]^T$,
$\boldsymbol{\theta}_{\rm R} = [\theta_1^r, \phi_1^r, \theta_2^r, \phi_2^r, \cdots, \theta_L^r, \phi_L^r]^T$,
$\boldsymbol{\theta}_{\rm T} = [\theta_1^t, \phi_1^t, \theta_2^t, \phi_2^t, \cdots, \theta_L^t, \phi_L^t]^T$,
$\mathbf{A}_{\rm R}(\boldsymbol{\theta}_{\rm R}) = [\mathbf{a}_{\rm R}(\theta_1^r, \phi_1^r), \mathbf{a}_{\rm R}(\theta_2^r, \phi_2^r), \cdots, \mathbf{a}_{\rm R}(\theta_L^r, \phi_L^r)]$,
$\mathbf{A}_{\rm T}(\boldsymbol{\theta}_{\rm T}) = [\mathbf{a}_{\rm T}(\theta_1^t, \phi_1^t), \mathbf{a}_{\rm T}(\theta_2^t, \phi_2^t), \cdots, \mathbf{a}_{\rm T}(\theta_L^t, \phi_L^t)]$.

From (7), we find that the channel $\mathbf{H}$ is related to path gains $\boldsymbol{\alpha}$ and $(\boldsymbol{\theta}_{\rm R}, \boldsymbol{\theta}_{\rm T})$. The angle-domain sparsity of the channel matrix, that is, the sparse channel estimation problem can be formulated as

$$\min_{\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\theta}_{\rm R}}, \widehat{\boldsymbol{\theta}_{\rm T}}} \quad \|\widehat{\boldsymbol{\alpha}}\|_0, \quad s.t. \left\|\mathbf{Y} - \mathbf{W}^H \hat{\mathbf{H}}\mathbf{X}\right\|_{\rm F} \leq \sigma \ , \tag{8}$$

where $\|\cdot\|_0$ denotes the l_0 norm. $(\hat{\boldsymbol{\theta}}_{\rm R}, \hat{\boldsymbol{\theta}}_{\rm T})$, $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{H}}$ are the estimate values corresponding to the original terms, and σ is the error tolerance parameter.

III. The Proposed Iterative Reweight Based Log-Sum Constraint Channel Estimation

A. Model optimization method

Equation (8) is an NP-hard problem, which means that the l_0-norm cannot find an efficient optimal solution in terms of computation. Alternative method such as l_1-norm can be utilized to replace the l_0-norm to achieve a sparse optimal solution. We consider log-sum item as sparsity-promoting functional. It has been proven that the log-sum penalty function has more sparsity-encouraging than the l_1-norm in [13]. Equation (8) can be written as

$$\begin{aligned}\min_{\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}} \quad & F(\boldsymbol{\alpha}) \triangleq \sum_{l=0}^{L}\log\left(\varepsilon+\frac{|\boldsymbol{\alpha}_l|^2}{\eta}\right)\\ & s.\ t.\ \left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_F \le \sigma,\end{aligned} \tag{9}$$

where ε and η are positive constants to ensures that the logarithmic function is well-defined. To realize an unconstrained optimization, we add a regularization parameter λ_0. Equation (9) can be expressed as

$$\begin{aligned}\min_{\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}} \quad G(\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}) &\triangleq \sum_{l=1}^{L}\log\left(\varepsilon+\frac{|\boldsymbol{\alpha}_l|^2}{\eta}\right)\\ &+\lambda_0\left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_F^2\\ &= L_0(\boldsymbol{\alpha})+\lambda_0\left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_F^2.\end{aligned} \tag{10}$$

where $L_0(\boldsymbol{\alpha}) = \sum_{l=1}^{L}\log\left(\varepsilon+\frac{|\boldsymbol{\alpha}_l|^2}{\eta}\right)$. We now utilize an iterative reweighted algorithm to solve the optimization (10). The main idea is to find a suitable surrogate function for (10). Obviously, the log-sum item can be replaced by

$$J(\boldsymbol{\alpha}|\hat{\boldsymbol{\alpha}}^{(t)}) \triangleq \sum_{l=1}^{L}\left(\frac{\varepsilon+\frac{|\boldsymbol{\alpha}_l|^2}{\eta}}{\varepsilon+\frac{\left|\hat{\boldsymbol{\alpha}}_l^{(t)}\right|^2}{\eta}}+\log\left(\varepsilon+\frac{\left|\hat{\boldsymbol{\alpha}}_l^{(t)}\right|^2}{\eta}\right)-1\right), \tag{11}$$

where $\hat{\boldsymbol{\alpha}}^{(t)} \triangleq \left[\hat{\alpha}_1^{(t)},\hat{\alpha}_2^{(t)},\cdots,\hat{\alpha}_L^{(t)}\right]^T$ is an estimate of $\boldsymbol{\alpha}$ at iteration t. It can be easily proven that $J(\boldsymbol{\alpha}|\hat{\boldsymbol{\alpha}}^t)-L_0(\boldsymbol{\alpha})\ge 0$, and the equality hold when $\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}^{(t)}$. As a result, the surrogate function for (10) can be expressed as

$$\mathrm{Z}(\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}) = J(\boldsymbol{\alpha}|\hat{\boldsymbol{\alpha}}^t)+\lambda_0\left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_F^2. \tag{12}$$

Consequently, the solution to (10) can be transformed into minimizing the surrogate function iteratively. Further, (12) is simplified as in [12] that

$$\mathbf{Z}^{(t)}(\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}) \triangleq \lambda_0^{-1}\boldsymbol{\alpha}^H\mathbf{K}^{(t)}\boldsymbol{\alpha}+\left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_F^2, \tag{13}$$

where $\mathbf{K}^{(t)}$ is given as

$$\mathbf{K}^{(t)} \triangleq diag\left(\frac{1}{\varepsilon+\frac{\left|\hat{\boldsymbol{\alpha}}_1^{(t)}\right|^2}{\eta}},\frac{1}{\varepsilon+\frac{\left|\hat{\boldsymbol{\alpha}}_2^{(t)}\right|^2}{\eta}},\cdots,\frac{1}{\varepsilon+\frac{\left|\hat{\boldsymbol{\alpha}}_L^{(t)}\right|^2}{\eta}}\right). \tag{14}$$

Then, we can optimize (13) with respect to the path gains $\boldsymbol{\alpha}$, that is

$$\min_{\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}} \mathbf{Z}^{(t)}(\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}) \triangleq \lambda_0^{-1}\boldsymbol{\alpha}^H\mathbf{K}^{(t)}\boldsymbol{\alpha}+\left\|\mathbf{Y}-\mathbf{W}^H\hat{\mathbf{H}}\mathbf{X}\right\|_{\mathrm{F}}^2. \tag{15}$$

We can ignore the superscript (t) of $\mathbf{Z}^{(t)}$ and $\mathbf{K}^{(t)}$ in (15), and simplify $\mathbf{A}_{\mathrm{R}}(\boldsymbol{\theta}_{\mathrm{R}})$ and $\mathbf{A}_{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{T}})$ with $\mathbf{A}_{\mathrm{R}}$ and $\mathbf{A}_{\mathrm{T}}$ respectively. The objective function $\mathbf{Z}$ can be expressed as

$$\begin{aligned}\mathbf{Z}(\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}) &= \lambda_0^{-1}\boldsymbol{\alpha}^H\boldsymbol{K}\boldsymbol{\alpha}\\ &\quad+\sum_{m=1}^{C_{\mathrm{X}}}\left\|\mathbf{y}_m-\mathbf{W}^H\mathbf{A}_{\mathrm{R}}diag(\boldsymbol{\alpha})\mathbf{A}_{\mathrm{T}}^H\mathbf{x}_m\right\|_2^2\\ &= \lambda_0^{-1}\boldsymbol{\alpha}^H\boldsymbol{K}\boldsymbol{\alpha}\\ &\quad+\sum_{m=1}^{C_{\mathrm{X}}}\left(\mathbf{y}_m-\mathbf{P}_m\boldsymbol{\alpha}\right)^H\left(\mathbf{y}_m-\mathbf{P}_m\boldsymbol{\alpha}\right)\\ &= \lambda_0^{-1}\boldsymbol{\alpha}^H\boldsymbol{K}\boldsymbol{\alpha}+\boldsymbol{\alpha}^H\left(\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{P}_m\right)\boldsymbol{\alpha}\\ &\quad-\boldsymbol{\alpha}^H\left(\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{y}_m\right)-\left(\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{y}_m^H\mathbf{P}_m\right)\boldsymbol{\alpha}\\ &\quad+\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{y}_m^H\mathbf{y}_m\quad,\end{aligned} \tag{16}$$

where $\mathbf{P}_m=\mathbf{W}^H\mathbf{A}_{\mathrm{R}}diag(\mathbf{A}_{\mathrm{T}}^H\mathbf{x}_m)$, then the partial derivative of (16) can be expressed as

$$\frac{\partial\mathbf{Z}(\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_T)}{\partial\boldsymbol{\alpha}} = \boldsymbol{\alpha}^H\left(\lambda_0^{-1}\mathbf{K}+\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{P}_m\right)-\left(\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{y}_m^H\mathbf{P}_m\right). \tag{17}$$

The optimal solution $\boldsymbol{\alpha}$ and corresponding minimum value of $\mathbf{Z}(\boldsymbol{\alpha},\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}})$ can be calculated when (17) is set to 0, we have

$$\boldsymbol{\alpha}_{opt}^{(t)}(\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}) = \left(\lambda_0^{-1}\mathbf{K}+\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{P}_m\right)^{-1}\left(\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{y}_m\right). \tag{18}$$

Combine (16) and (18), the optimal objective function can be expressed as

$$\begin{aligned}\mathbf{Z}_{opt}^{(t)}(\boldsymbol{\theta}_{\mathrm{R}},\boldsymbol{\theta}_{\mathrm{T}}) = -&\left(\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{y}_m\right)^H\left(\lambda_0^{-1}\mathbf{K}+\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{P}_m\right)^{-1}\\ &\cdot\left(\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{y}_m\right)+\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{y}_m^H\mathbf{y}_m\end{aligned}. \tag{19}$$

B. Iterative Reweight-based Log-Sum Constraint Channel Estimation

As mentioned earlier, the constrained optimization problem (8) has been replaced with an unconstrained angle optimization problem (19). Next, we only need to optimize the normalized spatial angles $\boldsymbol{\theta}_{\mathrm{R}}$ and $\boldsymbol{\theta}_{\mathrm{T}}$ to get the final result.

Specifically, at the tth iteration, the gradient descent method is used to search for the new estimates $\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(t+1)}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(t+1)}$, given as

$$\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(t+1)} = \hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(t)}-\xi\cdot\nabla_{\theta_{\mathrm{R}}}\mathbf{Z}_{opt}^{(t)}(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(t)},\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(t)}), \tag{20}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(t+1)} = \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(t)}-\xi\cdot\nabla_{\theta_{\mathrm{T}}}\mathbf{Z}_{opt}^{(t)}(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(t)},\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(t)}), \tag{21}$$

where ξ is the step length. With iterative searching, the estimates of $(\theta_{\mathrm{R}},\theta_{\mathrm{T}})$ can be closer to their actual positions, thus, high-precision channel estimation can be realized.

We assume $\mathbf{D}=\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{y}_m$,$\mathbf{B}=\sum_{m=1}^{C_{\mathrm{X}}}\mathbf{P}_m^H\mathbf{P}_m$ the gradients can be calculated as follows

$$\begin{aligned}\frac{\partial\mathbf{Z}_{opt}}{\partial\theta_{\mathrm{R},l}} &= -\frac{\partial\mathbf{D}^H}{\partial\theta_{\mathrm{R},l}}\mathbf{B}^{-1}\mathbf{D}-\mathbf{D}^H\frac{\partial\mathbf{B}^{-1}}{\partial\theta_{\mathrm{R},l}}\mathbf{D}-\mathbf{D}^H\mathbf{B}^{-1}\frac{\partial\mathbf{D}}{\partial\theta_{\mathrm{R},l}}\\ &= -\frac{\partial\mathbf{D}^H}{\partial\theta_{\mathrm{R},l}}\mathbf{B}^{-1}\mathbf{D}+\mathbf{D}^H\mathbf{B}^{-1}\frac{\partial\mathbf{B}}{\partial\theta_{\mathrm{R},l}}\mathbf{B}^{-1}\mathbf{D}\\ &\quad-\mathbf{D}^H\mathbf{B}^{-1}\frac{\partial\mathbf{D}}{\partial\theta_{\mathrm{R},l}},\end{aligned} \tag{22}$$

where $\frac{\partial \mathbf{D}^H}{\partial \theta_{\mathrm{R},l}} = \sum_{m=1}^{C_x} \frac{\partial \mathbf{P}_m^H}{\partial \theta_{\mathrm{R},l}} \mathbf{y}_m$,

$\frac{\partial \mathbf{B}}{\partial \theta_{\mathrm{R},l}} = \sum_{m=1}^{C_x} \left(\frac{\partial \mathbf{P}_m^H}{\partial \theta_{\mathrm{R},l}} \mathbf{P}_m + \mathbf{P}_m^H \frac{\partial \mathbf{P}_m}{\partial \theta_{\mathrm{R},l}} \right)$,

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

To reduce the computational complexity, the choice of initial $\left(\boldsymbol{\theta}_{\mathrm{R}}^{(0)}, \boldsymbol{\theta}_{\mathrm{T}}^{(0)} \right)$ becomes very important. SVD-based preconditioning can solve this problem effectively. Specifically, from (3) and (7), we have

$$\mathbf{Y} = \left(\mathbf{W}^H \mathbf{A}_{\mathrm{R}}(\boldsymbol{\theta}_{\mathrm{R}}) \right) diag(\boldsymbol{\alpha}) \left(\mathbf{X}^H \mathbf{A}_{\mathrm{T}}(\boldsymbol{\theta}_{\mathrm{T}}) \right)^H + \mathbf{N}_o. \quad (23)$$

Then, singular value decomposition can be applied to the matrix $\mathbf{Y}$, i.e.,$\mathbf{Y} = \mathbf{U} \sum \mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are singular matrix and subject to unit orthogonality, $\sum = diag\left(\sigma_1, \sigma_2, \ldots, \sigma_{\min(C_{\mathrm{X}}, C_{\mathrm{Y}})} \right)$ and its diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(C_{\mathrm{X}}, C_{\mathrm{Y}})} \geq 0$ are the singular values of $\mathbf{Y}$. The largest L singular values and their corresponding singular vectors can be approximately expressed as

$$\begin{aligned} \sigma_i &\approx |\alpha_{l_i}| \left\| \mathbf{W}^H \mathbf{a}_{\mathrm{R}}(\theta_{l_i}^r, \phi_{l_i}^r) \right\|_2 \left\| \mathbf{X}^H \mathbf{a}_{\mathrm{R}}(\theta_{l_i}^t, \phi_{l_i}^t) \right\|_2, \\ \mathbf{u}_i &\approx \mathbf{W}^H \mathbf{a}_{\mathrm{R}}(\theta_{l_i}^r, \phi_{l_i}^r) / \left\| \mathbf{W}^H \mathbf{a}_{\mathrm{R}}(\theta_{l_i}^t, \phi_{l_i}^t) \right\|_2, \\ \mathbf{v}_i &\approx \mathbf{X}^H \mathbf{a}_{\mathrm{T}}(\theta_{l_i}^r, \phi_{l_i}^r) / \left\| \mathbf{X}^H \mathbf{a}_{\mathrm{T}}(\theta_{l_i}^t, \phi_{l_i}^t) \right\|_2, \end{aligned} \quad (24)$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the ith column of $\mathbf{U}$ and $\mathbf{V}$, $i = 1, 2, \ldots L$,$\{l_1, l_2, \cdots, l_L\}$ is a permutation of $\{1, 2, \ldots L\}$. By utilizing this method, we can search for the nearest angle-domain grids to real AoAs/AoDs.

In the iterative objective function $\mathbf{Z}^{(t)}(\boldsymbol{\alpha}, \boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}})$, the dynamic regularization parameter λ_0 is updated with the iteration and controls the tradeoff between the sparsity and the data fitting error. Specifically, a smaller λ_0 can result in a sparse solution but cause an underestimation, while a larger λ_0 leads to a less sparse but better-fitting solution. Therefore, choosing λ_0 in the proposed algorithm can be more intelligent and updated by

$$d_e^{(t)} = \left\| \mathbf{Y} - \mathbf{W}^H \mathbf{A}_{\mathrm{R}}(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(t)}) diag(\hat{\boldsymbol{\alpha}}^{(t)}) \mathbf{A}_{\mathrm{T}}^H(\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(t)}) \mathbf{X} \right\|_F^2 \quad (25)$$

$$\lambda_0 = \min\left(\frac{c}{d_e^{(t)}}, \lambda_{\max} \right), \quad (26)$$

where c is a constant scaling factor, and $\lambda_{\max}$ is a threshold parameter.

IV. Simulation Results

In this section, we demonstrate the performance of the proposed algorithms for downlink channel estimation. Specifically, the model we apply is a mmWave massive MIMO system with hybrid precoding, with $L = 3$, $d = \lambda/2$, $N_{\mathrm{R}} = N_{\mathrm{T}} = 64/128$, $N_{\mathrm{R}}^{\mathrm{RF}} = N_{\mathrm{T}}^{\mathrm{RF}} = 4$. We use MATLAB 2016 to simulate the proposed algorithm on PC.In line-of-sight (LoS) the gain of the paths is normalized to 1, while the gain of non-line-of-sight (NLoS) paths is assumed to follow $CN(0.0.1)$. The directions of all paths of users are assumed to follow the IID uniform distribution within $[-\pi/2 \ , \pi/2]$. Each element of the transmitted pilots X satisfies $x_{i,j} = \sqrt{\rho/N_{\mathrm{T}}} e^{jw_{i,j}}$, where ρ is the transmitted power and $w_{i,j}$ is the random phase uniformly distributed in $[0 \ , 2\pi]$. We assume that the path gains follow Gaussian distribution. Within a certain range of signal-to-noise ratio (SNR), normalized mean square error (NMSE) is used as a standard to evaluate the performance of the algorithm. The reference algorithms involved in performance comparison are: Oracle LS scheme, conventional orthogonal matching pursuit (OMP)-based channel estimation [14] and iterative reweight-based super resolution channel estimation scheme [12].

We consider the ULA geometry in case 1 and generate Figs. 2 (a) and (b), which compare the performance of the proposed algorithm when changing the number of receiving and transmitting antennas under line-of-sight (LoS). Obviously, the proposed method shows excellent performance when the SNR becomes larger compared with the traditional algorithm. As the number of antennas increases, the performance of the proposed algorithm is better. Particularly, the NMSE value of the proposed algorithm about 9.3×10^{-3} smaller than the iterative reweight-based super resolution channel estimation scheme.

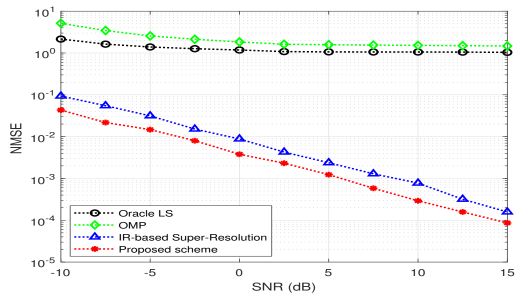
In case 2, we consider the performance of the proposed algorithm under non-line-of-sight (NLoS) transmission. The NMSE value difference is about 0.21 between the IR-based super-resolution algorithm and the proposed method in Fig. 3, that is 2.7×10^{-2} in Fig 4 of case 3 which consider a UPA geometry. In the above two cases, the proposed algorithm has excellent channel estimation performance. The UPA structure of case 3 is still carried out under the transmission conditions of LoS. Comparing Fig. 2 (a) and Fig.4, it can be seen that the proposed algorithm has better performance when adopting the ULA structure since the estimation errors of both azimuth and elevation angles contribute to the NMSE, under the same number of antennas and number of pilot overheads.

Based on the simulations and analysis above, it is easy to see that the proposed algorithm achieves much higher channel estimation accuracy,but has a higher computational complexity compared with traditional methods such as the OMP-based channel estimation scheme [14].

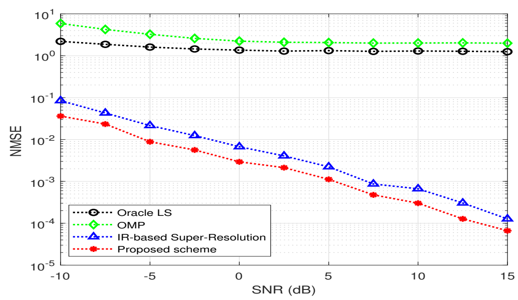
TABLE I
THE PROPOSED ALGORITHM

Input:$\mathbf{X}$, $\mathbf{Y}$, $\mathbf{W}$, L_{init}, α_{th}, ζ_{th}.
Output: Coarse AoAs/AoDs estimates $(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(0)}, \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(0)})$
Estimated AoAs/AoDs $(\hat{\boldsymbol{\theta}}_{\mathrm{R}}, \hat{\boldsymbol{\theta}}_{\mathrm{T}})$
and path gains of all paths.

1:$[\mathbf{U}, \sum, \mathbf{V}] = \mathrm{SVD}(\mathbf{Y})$.
2:**for** $i = 1, 2, \ldots L_{init}$
$(\hat{\theta}_i^{r(0)}, \phi_i^{r(0)}) = \arg\max \ \mathbf{u}_i \mathbf{W}^H \mathbf{a}_{\mathrm{R}}$
$(\hat{\theta}_i^{t(0)}, \phi_i^{t(0)}) = \arg\max \ \mathbf{v}_i \mathbf{W}^H \mathbf{a}_{\mathrm{R}}$.
3:**end for**
4:$\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(0)} = [\hat{\theta}_1^{r(0)}, \phi_1^{r(0)}, \hat{\theta}_2^{r(0)}, \phi_2^{r(0)}, \ldots, \hat{\theta}_{L_{init}}^{r(0)}, \phi_{L_{init}}^{r(0)}]$
$\hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(0)} = [\hat{\theta}_1^{t(0)}, \phi_1^{t(0)}, \hat{\theta}_2^{t(0)}, \phi_2^{t(0)}, \ldots, \hat{\theta}_{L_{init}}^{t(0)}, \phi_{L_{init}}^{t(0)}]$.
5:Initialize $\hat{\boldsymbol{\alpha}}^{(0)} = \boldsymbol{\alpha}_{opt}(\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(0)}, \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(0)})$ according to (17).
6:**Repeat**
7: Update λ by (24).
8: Construct $\mathbf{Z}_{opt}^{(t)}(\boldsymbol{\theta}_{\mathrm{R}}, \boldsymbol{\theta}_{\mathrm{T}})$ by (18).
9: Calculate $\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(t+1)}$,$\hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(t+1)}$ by (19), (20).
10: Estimate the path gains $\hat{\alpha}^{(t+1)}$ by (17).
11: If $\hat{\alpha}_l^{(t+1)} < \alpha_{th}$, prune path l.
12:**until** $L^{(t)} = L^{(t+1)}$ and $\left\| \hat{\boldsymbol{\alpha}}^{(t+1)} - \hat{\boldsymbol{\alpha}}^{(t)} \right\|_2 < \zeta_{th}$.
13:$\hat{\boldsymbol{\theta}}_{\mathrm{R}} = \hat{\boldsymbol{\theta}}_{\mathrm{R}}^{(last)}$, $\hat{\boldsymbol{\theta}}_{\mathrm{T}} = \hat{\boldsymbol{\theta}}_{\mathrm{T}}^{(last)}$, $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}^{(last)}$.



(a) $N_{\mathrm{T}} = N_{\mathrm{R}} = 64$



(b) $N_{\mathrm{T}} = N_{\mathrm{R}} = 128$

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

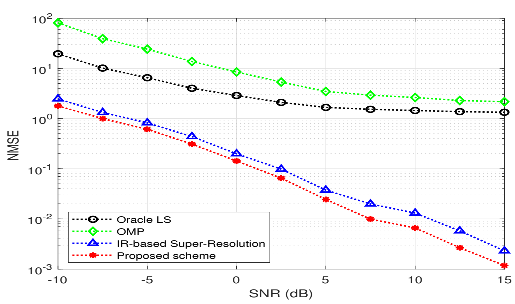


Fig. 3. NMSE performance comparison of different channel estimation schemes (NLoS)

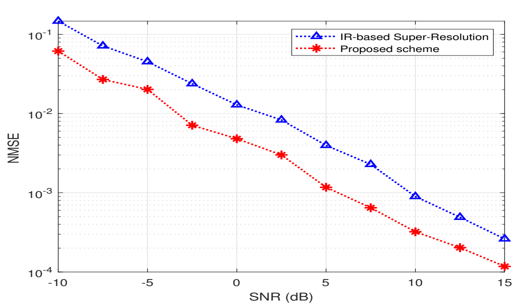


Fig. 4. NMSE performance comparison of different channel estimation schemes (UPA)

V. Conclusion

In this paper, we proposed a new iterative reweight-based log-sum constraint channel estimation scheme for mmWave massive MIMO with hybrid precoding. Specifically, we converted the channel estimation problem into a sparse optimization problem. The proposed new objective function consists of two parts: the sparse weighted sum and the data fitting error. We set a flexible regularization parameter to adjust the tradeoff between these two parts. For the new objective function, we utilized the traditional gradient descent method to iteratively move the coarse on-grid points to their actual off-grid positions. Moreover, a dynamic regularization parameter was applied to control the tradeoff between the channel sparsity and the data fitting error. The simulation results show that the proposed channel estimation algorithm achieves more superiority than several existing compressed sensing algorithms in terms of accuracy. Angle domain channel estimation is very important for mmWave massive MIMO. More accurate angle estimation is a practical method for achieving higher spectral efficiency. Further developments of the proposed approach can involve high mobility scenarios or other schemes to reduce complexity.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61501530 and Grant 61803396. Henan Educational Committee under Grant 21A510015.

References

[1] H. Huang *et al.*, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wirel. Commun.*, vol. 27, no. 1, pp. 214–222, 2020.

[2] J. Jin, C. Xiao, W. Chen, S. Member, and Y. Wu, "Channel-Statistics-Based Hybrid Precoding for Millimeter-Wave MIMO Systems With Dynamic Subarrays," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 3991-4003, 2019.

[3] D. Fan *et al.*, "Angle Domain Channel Estimation in Hybrid Millimeter Wave Massive MIMO Systems," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 12, pp. 8165–8179, 2018.

[4] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter Wave Channel Estimation via Exploiting Joint Sparse and Low-Rank Structures," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 2, pp. 1123–1133, 2018.

[5] M. J. Azizipour, K. Mohamed-Pour," Compressed channel estimation for FDD massive MIMO systems without prior knowledge of sparse channel model." *IET Commun.*, vol. 13, no. 6, pp. 657-663, 2018.

[6] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, "Wideband MIMO Channel Estimation for Hybrid Beamforming Millimeter Wave Systems via Random Spatial Sampling," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 5, pp. 1136–1150, 2019.

[7] M. Wang, F. Gao, N. Shlezinger, M. F. Flanagan, and Y. C. Eldar, "A block sparsity based estimator for mmWave Massive MIMO channels with beam squint," *IEEE Trans. Signal Process.*, vol. 68, pp. 49–64, 2020.

[8] L. Ge, Y. Zhang, G. Chen, and J. Tong, "Compression-Based LMMSE Channel Estimation with Adaptive Sparsity for Massive MIMO in 5G Systems," *IEEE Syst. J.*, vol. 13, no. 4, pp. 3847–3857, 2019.

[9] S. Rao, A. Mezghani, and A. Lee Swindlehurst, "Channel Estimation in One-Bit Massive MIMO Systems: Angular Versus Unstructured Models," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 5, pp. 1017–1031, 2019.

[10] Z. Xiao, P. Xia, and X. G. Xia, "Channel Estimation and Hybrid Precoding for Millimeter-Wave MIMO Systems: A Low-Complexity Overall Solution," *IEEE Access,* vol. 5, no. c, pp. 16100–16110, 2017.

[11] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse Channel Estimation and Hybrid Precoding Using Deep Learning for Millimeter Wave Massive MIMO," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2838–2849, 2020.

[12] C. Hu, "Super-Resolution Channel Estimation for MmWave Massive MIMO With Hybrid Precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, 2018.

[13] F. Wang, J. Fang, and H. Li, "Prior knowledge aided super-resolution line spectral estimation: An iterative reweighted algorithm," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process.- Proc.*, pp. 3296–3300, 2017.

[14] X. Gao, L. Dai, S. Zhou, A. M. Sayeed, and L. Hanzo, "Beamspace channel estimation for wideband millimeter-wave MIMO with lens antenna array," *IEEE Int. Conf. Commun.*, vol. 2018-May, pp. 1–6, 2018.