



# Article Low-Complexity Joint Channel Estimation for Multi-User mmWave Massive MIMO Systems

Jianhe Du<sup>1,\*</sup>, Jiaqi Li<sup>1</sup>, Jing He<sup>1</sup>, Yalin Guan<sup>1</sup> and Heyun Lin<sup>2</sup>

- <sup>1</sup> School of Information and Communication Engineering, Communication University of China, Beijing 100024, China; llijiaqi1@cuc.edu.cn (J.L.); hejing@cuc.edu.cn (J.H.); ylguan@cuc.edu.cn (Y.G.)
- <sup>2</sup> Power Dispatching Control Center, Guangxi Power Grid, Nanning 530023, China; forlan@bupt.edu.cn
- \* Correspondence: dujianhe@cuc.edu.cn

Received: 15 January 2020; Accepted: 5 February 2020; Published: 9 February 2020



**Abstract:** For multi-user millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) systems, the precise acquisition of channel state information (CSI) is a huge challenge. With the increase of the number of antennas at the base station (BS), the traditional channel estimation techniques encounter the problems of pilot training overhead and computational complexity increasing dramatically. In this paper, we develop a step-length optimization-based joint iterative scheme for multi-user mmWave massive MIMO systems to improve channel estimation performance. The proposed estimation algorithm provides the BS with full knowledge of all channel parameters involved in up- and down-links. Compared with existing algorithms, the proposed algorithm has higher channel estimation accuracy with low complexity. Moreover, the proposed scheme performs well even with a small number of training sequences and a large number of users. Simulation results are shown to demonstrate the performance of the proposed channel estimation algorithm.

Keywords: multi-user; mmWave; massive MIMO; channel estimation; low complexity

# 1. Introduction

Millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) has been generally recognized as a strong potential key technique for enhancing quality in the future wireless communication field [1]. The demand for higher speed data transmission, shorter delays, better user experience, and denser networks is increasing with the rapid development of wireless services such as virtual reality, multimedia, and the Internet of Things [2]. The mmWave band has huge unexploited spectrum resources, which can overcome the spectrum congestion of standard wireless frequency bands and achieve an orders of magnitude increase in spectral efficiency to meet these requirements [3,4]. The work in [5] expounded that the mmWave band can greatly speed up the data transmission rate. Additionally, massive MIMO refers to adding a number of antennas at the base station (BS), which can overcome the effects of fading, average out the dominant path-loss, and make up for the short wavelength of mmWave [6–8], and the isolation between radiating elements is a very important issue [9,10]. The combination of mmWave and massive MIMO can be used to meet increasing demand in data traffic [11,12].

Many studies have been devoted to the research of mmWave massive MIMO multi-user systems in recent years. Compared with conventional point-to-point MIMO, multi-user MIMO with simplified resource allocation has greater advantages [13,14]. The work [15] developed blind multi-user detection based on the model of [16], which can approach the lower bound of performance under certain conditions. The work in [17] summarized the benefits, challenges, and potential solutions of mmWave massive MIMO in cellular networks.

In order to achieve the theoretical performance gain and potential benefits of mmWave massive MIMO systems, accurate channel state information (CSI) is essential, which is also regarded as one of the main difficulties in mmWave communication [18,19]. However, due to the difference in channel characteristics compared to traditional sub-6 GHz, the smaller number of RF chains, and overwhelming overhead in downlink training and uplink channel feedback, accurate acquisition of CSI is challenging [20–22]. Various novel channel estimation approaches have been recently proposed for mmWave massive MIMO [23–27]. In particular, under the assumption of the BS antennas being highly correlated, the work in [23] proposed an antenna grouping method to reduce feedback overhead. By utilizing the correlation among users, the work in [24] provided a joint CSI acquisition proposal based on a low-rank matrix. To reduce errors caused by discretization and solve the line spectral estimation problem, the work in [25] designed an iterative reweighted approach for joint dictionary parameter training and sparse signal recovery, but it had the limitation that only two parameters could be estimated. The work in [26] extended the suggestion in [25] to two dimensions and represented a super-resolution channel estimation scheme based on iterative reweighting in a point-to-point scenario. It is regrettable that such a solution in [26] had the problem of slow estimation. Considering the problem of channel estimation for mmWave MIMO systems under a transmitter impairment model, a new algorithm based on Bayesian compressive sensing (BCS) was introduced in [27]. However, the assumption of BCS was relatively harsh, and the accuracy of estimation was limited.

To obtain CSI more accurately and faster in multi-user mmWave massive MIMO, we develop a step-length optimization-based joint iterative channel estimation scheme in this paper. Our main contributions are outlined as follows:

1. For multi-user mmWave massive MIMO systems, we propose a novel joint estimation scheme, which can accurately estimate the azimuth angles of departure or arrival (AoDs/AoAs) and joint gain at the BS. Specifically, the proposed algorithm does not need to estimate the channel on the user side, and all channel parameters can be estimated at the BS, which is different from the traditional channel estimation scheme.

2. We combine the advantages of the Newton method and gradient descent method and derive an iterative step-length optimization approach to reduce computational complexity. It is confirmed that the proposed scheme has low complexity without losing the accuracy of channel estimation.

3. Compared with the traditional singular-value decomposition (SVD) algorithm and BCS algorithm under different situations, the simulation results show that the proposed algorithm achieves better channel estimation performance with fewer training data blocks.

The rest of this paper is arranged as follows. Section 2 introduces the considered multi-user mmWave massive MIMO model. Section 3 formulates the proposed joint iterative channel estimation algorithm, which accounts for the preprocessing, the choice of descent direction in the iteration, the optimization of the step-length, and the method for separating the downlink channel. Some simulation results are provided to illustrate the performance of estimation in Section 4, and Section 5 concludes this paper.

Notation: In this paper, vectors and matrices are denoted by boldface lowercase and uppercase symbols, respectively;  $\mathbf{A}^T$ ,  $\mathbf{A}^H$ ,  $\mathbf{A}^{-1}$ , and  $\mathbf{A}^+$  correspond to the transpose, conjugate transpose, inverse, and Moore–Penrose pseudo-inverse of the matrix  $\mathbf{A}$ ; *vec* ( $\mathbf{A}$ ) is the vector stacked by the columns of the matrix  $\mathbf{A}$ ;  $\otimes$  is the Kronecker product; *diag*( $\mathbf{x}$ ) denotes the diagonal matrix with the vector  $\mathbf{x}$  on its diagonal; *rank*( $\mathbf{A}$ ) is the rank of  $\mathbf{A}$ ;  $\|\mathbf{A}\|_F$ ,  $\|\mathbf{h}\|_0$ , and  $\|\mathbf{h}\|_2$  represent the Frobenius norm of  $\mathbf{A}$ , the  $\ell_0$ -norm, and Euclidean norm of the vector  $\mathbf{h}$ , respectively;  $\mathbf{I}_N$  is the  $N \times N$  identity matrix; the sets of real and complex numbers are represented by  $\mathbb{R}$  and  $\mathbb{C}$ , respectively.

#### 2. System Model

We considered a common mmWave massive MIMO system with hybrid precoding as illustrated in Figure 1, where the number of antennas at the BS is  $N_{BS}$ ; but, only  $N_{RF}$  RF chains with  $N_{BS} > N_{RF}$  to support K user equipment (UE), and each UE has a single receive antenna [18,28]. In the communication process, it is assumed that there is only one transmission path via  $N_S$  data streams between BS and each UE.



Figure 1. Block diagram of the multi-user mmWave massive MIMO system.

Let  $m_T = 1, \dots, M_T$  and  $m_R = 1, \dots, M_R$  be the time slot of the transmitter and receiver, respectively. The BS pilots  $\mathbf{S}_{m_T,m_R} \in \mathbb{C}^{N_s \times L}$  represent the signal transmitted in the  $(m_T, m_R)^{\text{th}}$  time slot, where *L* is the number of different pilot sequences, and its design meets the orthogonality condition  $\mathbf{S}_{m_T,m_R} \cdot \mathbf{S}_{m_T,m_R}^{\text{H}} = \mathbf{I}_{N_S \times N_S}$  [29,30]. Then, the signal  $\mathbf{y}_k \in \mathbb{C}^{1 \times L}$  received by the *k*<sup>th</sup> user can be given by:

$$\mathbf{y}_k = \mathbf{h}_D^k \mathbf{F}_{T,m_T} \mathbf{S}_{m_T,m_R} + n_{m_T,m_R}^{MS} \tag{1}$$

where  $\mathbf{F}_{T,m_T} = [\mathbf{F}_{RF}]_{T,m_T} [\mathbf{F}_{BB}]_{T,m_T} \in \mathbb{C}^{N_{BS} \times N_S}$  is the transmitted hybrid precoding matrix in the  $m_T$ <sup>th</sup> time slot,  $\mathbf{F}_{BB} \in \mathbb{C}^{N_{RF} \times N_S}$  is the baseband beamformer followed by the RF beamformer  $\mathbf{F}_{RF} \in \mathbb{C}^{N_{BS} \times N_{RF}}$ ,  $n_{m_T,m_R}^{MS} \in \mathbb{C}^{1 \times L}$  is the received independent and identically distributed additive white Gaussian noise elements having zero mean and the variance  $\sigma_k^2$ , and  $\mathbf{h}_D^k \in \mathbb{C}^{1 \times N_{BS}}$  is the downlink channel vector, and by defining  $\theta_k \triangleq d \sin \delta_{T,k} / \lambda$  as the normalized spatial angle [31], it can be expressed as:

$$\mathbf{h}_{D}^{k} = \alpha_{k} \mathbf{a}^{H} \left( \theta_{k} \right) \tag{2}$$

where  $\alpha_k$ , d, and  $\lambda$  denote the propagation gain of the  $k^{\text{th}}$  downlink path and the antenna spacing at the BS and carrier wavelength, respectively.  $d = \frac{1}{2}\lambda$  is widely adopted here [32].  $\delta_{T,k} \in [0, 2\pi]$  is the AoD of the  $k^{\text{th}}$  path. **a** ( $\theta_k$ )  $\in \mathbb{C}^{N_{BS} \times 1}$  is the steering vector at the transmitter of the  $k^{\text{th}}$  path. In this work, we considered the typical uniform linear arrays (ULAs) [33]; **a** ( $\theta_k$ ) is given by:

$$\mathbf{a}\left(\theta_{k}\right) = \left[1, e^{j2\pi\theta_{k}}, \cdots, e^{j2\pi(N_{BS}-1)\theta_{k}}\right]^{T}.$$
(3)

Denote  $\phi_k \triangleq d \sin \delta_{R,k} / \lambda$  where  $\delta_{R,k}$  is the AoA of the  $k^{\text{th}}$  path.  $\phi_k$  is the normalized spatial angle the same as  $\theta_k$ . The uplink channel vector  $\mathbf{h}_U^k \in \mathbb{C}^{N_{BS} \times 1}$  can be expressed as:

$$\mathbf{h}_{U}^{k} = \beta_{k} \mathbf{b} \left( \phi_{k} \right) \tag{4}$$

where  $\beta_k$  is the propagation gain of the  $k^{\text{th}}$  uplink path and  $\mathbf{b}(\phi_k) \in \mathbb{C}^{N_{BS} \times 1}$  is the steering vector at the receiver of the  $k^{\text{th}}$  path. For typical ULAs,  $\mathbf{b}(\phi_k)$  is given by:

$$\mathbf{b}\left(\phi_{k}\right) = \left[1, e^{j2\pi\phi_{k}}, \cdots, e^{j2\pi(N_{BS}-1)\phi_{k}}\right]^{T}.$$
(5)

To reduce the overhead of downlink channel training and uplink channel feedback, we propose a joint estimation scheme, that is the user does not process the received signal and directly feeds it back to the BS [24]. For all *K* users, the comprehensive signal  $\mathbf{X}_{m_T,m_R}^{BS} \in \mathbb{C}^{N_S \times L}$  received by the BS is:

$$\mathbf{X}_{m_{T},m_{R}}^{BS} = \mathbf{W}_{R,m_{R}}^{H} \sum_{k=1}^{K} \mathbf{h}_{U}^{k} \mathbf{y}_{k}$$

$$= \mathbf{W}_{R,m_{R}}^{H} \sum_{k=1}^{K} \beta_{k} \mathbf{b} (\phi_{k}) \alpha_{k} \mathbf{a}^{H} (\theta_{k}) \mathbf{F}_{T,m_{T}} \mathbf{S}_{m_{T},m_{R}} + \mathbf{N}_{m_{T},m_{R}}^{BS}$$
(6)

where  $K \ll N_{BS}$  is the number of complex paths,  $\mathbf{W}_{R,m_R} = [\mathbf{W}_{RF}]_{R,m_R} [\mathbf{W}_{BB}]_{R,m_R} \in \mathbb{C}^{N_{BS} \times N_S}$  is the received hybrid combining matrix in the  $m_R^{\text{th}}$  time slot,  $\mathbf{W}_{RF}$  and  $\mathbf{W}_{BB}$  are baseband combiners, and  $\mathbf{N}_{m_T,m_R}^{BS} \in \mathbb{C}^{N_S \times L}$  is the noise received by the BS after combining. By defining  $r_k = \alpha_k \beta_k$  and synthesizing all *K* users, we can further obtain:

$$\mathbf{X}_{m_T,m_R}^{BS} = \mathbf{W}_{R,m_R}^H \mathbf{B}\left(\boldsymbol{\phi}\right) diag\left(\mathbf{r}\right) \mathbf{A}^H\left(\boldsymbol{\theta}\right) \mathbf{F}_{T,m_T} \mathbf{S}_{m_T,m_R} + \mathbf{N}_{m_T,m_R}^{BS}$$
(7)

where  $\boldsymbol{\phi} = [\phi_1, \phi_2, \cdots, \phi_k]^T \in \mathbb{C}^{K \times 1}, \boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_k]^T \in \mathbb{C}^{K \times 1}, \mathbf{r} = [r_1, r_2, \cdots, r_k]^T \in \mathbb{C}^{K \times 1},$ the steering matrix  $\mathbf{B}(\boldsymbol{\phi}) \in \mathbb{C}^{N_{BS} \times K}$  and  $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{C}^{N_{BS} \times K}$  are given by:

$$\mathbf{B}(\boldsymbol{\phi}) = [\mathbf{b}(\phi_1), \mathbf{b}(\phi_2), \cdots, \mathbf{b}(\phi_k)]$$
(8)

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_k)].$$
(9)

We further process the signal  $\mathbf{X}_{m_T,m_R}^{BS}$  in (7) to make it multiply right by  $\mathbf{S}_{m_T,m_R}^H$ ; we have:

$$\mathbf{X}_{m_T,m_R} = \mathbf{X}_{m_T,m_R}^{BS} \mathbf{S}_{m_T,m_R}^H = \mathbf{W}_{R,m_R}^H \mathbf{B}\left(\boldsymbol{\phi}\right) diag\left(\mathbf{r}\right) \mathbf{A}^H\left(\boldsymbol{\theta}\right) \mathbf{F}_{T,m_T} + \mathbf{N}_{m_T,m_R}.$$
 (10)

Next, we start to collect and arrange the signal  $X_{m_T,m_R}$  in (10). By collecting the pilots in the  $M_T$  time slots of the transmitter, we can get:

$$\mathbf{X}_{m_R} = \begin{bmatrix} \mathbf{X}_{1,m_R}, \mathbf{X}_{2,m_R}, \cdots, \mathbf{X}_{M_T,m_R} \end{bmatrix}$$
  
=  $\mathbf{W}_{R,m_P}^H \mathbf{B}\left(\boldsymbol{\phi}\right) diag\left(\mathbf{r}\right) \mathbf{A}^H\left(\boldsymbol{\theta}\right) \mathbf{F}_T + \mathbf{N}_{m_R}$  (11)

where  $\mathbf{F}_T = [\mathbf{F}_{T,1}, \cdots, \mathbf{F}_{T,M_T}] \in \mathbb{C}^{N_{BS} \times N_S M_T}$ ,  $\mathbf{N}_{m_R} = [\mathbf{N}_{1,m_R}, \cdots, \mathbf{N}_{M_T,m_R}] \in \mathbb{C}^{N_S \times N_S M_T}$ . By dealing with the signal  $\mathbf{X}_{m_R}$  in the  $M_R$  time slots of the receiver, we have after arranging:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1, \mathbf{X}_1, \cdots, \mathbf{X}_{M_R} \end{bmatrix}^T$$
  
=  $\mathbf{W}_R \mathbf{B}(\boldsymbol{\phi}) \operatorname{diag}(\mathbf{r}) \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{F}_T + \mathbf{N}$  (12)

where  $\mathbf{W}_{R} \in \mathbb{C}^{N_{S}M_{R} \times N_{BS}}$  and  $\mathbf{N} \in \mathbb{C}^{N_{S}M_{R} \times N_{S}M_{T}}$  are expressed as:

$$\mathbf{W}_{R} = \begin{bmatrix} \mathbf{W}_{R,1}^{H} \\ \vdots \\ \mathbf{W}_{R,M_{R}}^{H} \end{bmatrix}, \ \mathbf{N} = \begin{bmatrix} \mathbf{N}_{1} \\ \vdots \\ \mathbf{N}_{M_{R}} \end{bmatrix}.$$
(13)

Now, the signal received by the BS in (10) is written in a more compact form as:

$$\mathbf{X} = \mathbf{W}_{R}\mathbf{B}\left(\boldsymbol{\phi}\right) diag\left(\mathbf{r}\right)\mathbf{A}^{H}\left(\boldsymbol{\theta}\right)\mathbf{F}_{T} + \mathbf{N}$$
(14)

where  $\mathbf{X} \in \mathbb{C}^{N_S M_R \times N_S M_T}$ ; the combined channel matrix  $\mathbf{H} \in \mathbb{C}^{N_{BS} \times N_{BS}}$  can be written as:

$$\mathbf{H} = \mathbf{B}\left(\boldsymbol{\phi}\right) diag\left(\mathbf{r}\right) \mathbf{A}^{H}\left(\boldsymbol{\theta}\right).$$
(15)

# 3. Joint Iterative Channel Estimation Scheme

In this section, we firstly transform the above channel estimation problem into a new algorithm form. Then, we adopt the SVD preconditioning proposal to reduce the computational complexity. To improve algorithm performance, we give an angle estimation optimization scheme by optimizing the step-length and using a combination of the gradient descent algorithm and Newton method. Finally, we propose a scheme for successfully separating the downlink channel from the combined channel matrix.

## 3.1. Algorithm Formulation

The estimation of the combined channel matrix **H** in (15) is equivalent to the estimation of combined path gain **r** and the two normalized spatial angles  $\theta$  and  $\phi$  for all paths. Due to the sparseness of the combined channel matrix **H** in the angle domain [34], such a problem can be easily expressed as:

$$\min_{\mathbf{r},\boldsymbol{\theta},\boldsymbol{\phi}} \|\mathbf{r}\|_{0}$$

$$s.t. \|\mathbf{X} - \mathbf{W}_{R}\mathbf{H}\mathbf{F}_{T}\|_{F} \leq \gamma$$
(16)

where  $\|\mathbf{r}\|_0$  is the number of non-zero components of  $\mathbf{r}$ , which also represents the number of estimated paths *K*, and  $\gamma$  is the error tolerance parameter related to noise.

Considering the low computational efficiency of finding the optimal solution in (16), we propose to replace the  $\ell_0$ -norm with the use of the logarithmic sum function [25,35]:

$$\min_{\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi}} P(\mathbf{r}) = \sum_{k=1}^{K} \log(|r_k|^2 + \xi)$$
s.t.  $\|\mathbf{X} - \mathbf{W}_R \mathbf{H} \mathbf{F}_T\|_F \leq \gamma$ 
(17)

where  $r_k$  is the  $k^{\text{th}}$  element of the gain **r** and  $\xi > 0$  is a positive parameter to ensure that the definition of the log-sum function is valid. By adding a data fitting parameter  $\mu > 0$ , the difficulty in (17) can be formulated as an unconstrained optimization problem, which yields the following optimization:

$$\min_{\mathbf{r},\boldsymbol{\theta},\boldsymbol{\phi}} L\left(\mathbf{r},\boldsymbol{\theta},\boldsymbol{\phi}\right) = P\left(\mathbf{r}\right) + \mu \left\|\mathbf{X} - \mathbf{W}_{R}\mathbf{H}\mathbf{F}_{T}\right\|_{F}^{2}.$$
(18)

Due to the need to ensure the monotonically decreasing characteristic, we introduce a suitable iterative function instead of the log-sum function [26,36]:

$$\min_{\mathbf{r},\boldsymbol{\theta},\boldsymbol{\phi}} Z^{(t)}\left(\mathbf{r},\boldsymbol{\theta},\boldsymbol{\phi}\right) = \mu^{-1} \mathbf{r}^{H} \mathbf{G}^{(t)} \mathbf{r} + \|\mathbf{X} - \mathbf{W}_{R} \mathbf{H} \mathbf{F}_{T}\|_{F}^{2}$$
(19)

where *t* is the number of iterations and  $\mathbf{G}^{(t)}$  is a diagonal matrix, which is defined as:

$$\mathbf{G}^{(t)} \triangleq diag\left[\varphi_1, \varphi_2, \cdots, \varphi_k\right] \tag{20}$$

where  $\varphi_k = \frac{1}{\left|\tilde{r}_k^{(t)}\right| + \xi}$ ,  $\tilde{r}_k^{(t)}$  is the path gain estimate of the  $k^{\text{th}}$  user at the  $t^{\text{th}}$  iteration. The minimization of the log-sum function  $L(\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi})$  in (18) is equivalent to the minimization of the surrogate function

 $Z^{(t)}(\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi})$  in (19), which is proven as follows. To get better estimates  $\mathbf{\tilde{r}}^{(t+1)}, \mathbf{\tilde{\theta}}^{(t+1)}, \mathbf{\tilde{\phi}}^{(t+1)}$  in the *t*<sup>th</sup> iteration and ensure the convergence of the function, the following inequality can be shown:

$$Z^{(t)}\left(\tilde{\mathbf{r}}^{(t+1)}, \left(\tilde{\boldsymbol{\theta}}^{(t+1)}, \tilde{\boldsymbol{\phi}}^{(t+1)}\right) \middle| \tilde{\mathbf{r}}^{(t)}\right) \leqslant Z^{(t)}\left(\tilde{\mathbf{r}}^{(t)}, \left(\tilde{\boldsymbol{\theta}}^{(t)}, \tilde{\boldsymbol{\phi}}^{(t)}\right) \middle| \tilde{\mathbf{r}}^{(t)}\right).$$
(21)

Then, we can get:

$$L\left(\tilde{\mathbf{r}}^{(t+1)}, \tilde{\boldsymbol{\theta}}^{(t+1)}, \tilde{\boldsymbol{\phi}}^{(t+1)}\right) - \mu Z^{(t)}\left(\tilde{\mathbf{r}}^{(t+1)}, \left(\tilde{\boldsymbol{\theta}}^{(t+1)}, \tilde{\boldsymbol{\phi}}^{(t+1)}\right) \middle| \tilde{\mathbf{r}}^{(t+1)}\right) = \sum_{k=1}^{K} \left[\log\left(\left|\tilde{r}_{k}^{(t+1)}\right|^{2} + \xi\right) - \left|\tilde{r}_{k}^{(t+1)}\right|^{2} \varphi_{k}\right].$$
(22)

Since the maximum value is reached at  $\mathbf{\tilde{r}}^{(t+1)} = \mathbf{\tilde{r}}^{(t)}$ , we have:

$$\sum_{k=1}^{K} \left[ \log\left( \left| \tilde{r}_{k}^{(t+1)} \right|^{2} + \xi \right) - \left| \tilde{r}_{k}^{(t+1)} \right|^{2} \varphi_{k} \right] \leqslant \sum_{k=1}^{K} \left[ \log\left( \left| \tilde{r}_{k}^{(t+1)} \right|^{2} + \xi \right) - \left| \tilde{r}_{k}^{(t)} \right|^{2} \varphi_{k} \right].$$
(23)

Combining (21) and (22), we can finally get:

$$L\left(\tilde{\mathbf{r}}^{(t+1)}, \tilde{\boldsymbol{\theta}}^{(t+1)}, \tilde{\boldsymbol{\phi}}^{(t+1)}\right) \leqslant L\left(\tilde{\mathbf{r}}^{(t)}, \tilde{\boldsymbol{\theta}}^{(t)}, \tilde{\boldsymbol{\phi}}^{(t)}\right)$$
(24)

which explains the correctness of the above assumption.

By optimizing the path gain **r** in (19) through the derivative method [26], we can obtain the optimal point of **r** and the corresponding optimal value of  $Z^{(t)}$ , that is,

$$\mathbf{r}_{pe}^{(t)}(\boldsymbol{\theta}, \boldsymbol{\phi}) = (\mu^{-1} \mathbf{G}^{(t)} + \sum_{i=1}^{N_S M_T} \mathbf{Q}_i^H \mathbf{Q}_i)^{-1} (\sum_{i=1}^{N_S M_T} \mathbf{Q}_i^H \mathbf{x}_i)$$
(25)

$$Z_{pe}^{(t)}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_{i=1}^{N_s M_T} \mathbf{x}_i^H \mathbf{x}_i - (\sum_{i=1}^{N_s M_T} \mathbf{Q}_i^H \mathbf{x}_i)^H (\mu^{-1} \mathbf{G}^{(t)} + \sum_{i=1}^{N_s M_T} \mathbf{Q}_i^H \mathbf{Q}_i)^{-1} (\sum_{i=1}^{N_s M_T} \mathbf{Q}_i^H \mathbf{x}_i)$$
(26)

where  $\mathbf{x}_i$  is the column vector of the signal **X** received by the BS,  $\mathbf{Q}_i = \mathbf{W}_R \mathbf{B}(\boldsymbol{\phi}) diag(\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{f}_T)$ , and  $\mathbf{f}_T$  is the column vector of the hybrid precoding matrix  $\mathbf{F}_T$ . After that, we only need to consider the optimization of the normalized spatial angles  $\boldsymbol{\theta}$  and  $\boldsymbol{\phi}$  in (26), which will be discussed in the next subsection.

## 3.2. Initial Angle Preprocessing

To optimize the spatial angle, we propose a preconditioning algorithm based on SVD [37,38] to select the initial value of the spatial angle effectively. Compared with using all  $N_{MS}N_{BS}$  angle domain grids as the initial candidate value, the proposed algorithm can rapidly find the angle domain grids closest to the true orientation angle, which can significantly reduce the computational complexity of the subsequent algorithm. Specifically, we apply SVD to the matrix **X** in (14) and get  $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ , where **U** and **V** are unitary matrices that satisfy  $\mathbf{U}^H\mathbf{U} = \mathbf{I}_{N_SM_R \times N_SM_R}$  and  $\mathbf{V}^H\mathbf{V} = \mathbf{I}_{N_SM_T \times N_SM_T}$ , respectively,  $\mathbf{\Sigma} = diag\left(\kappa_1, \kappa_2, \cdots, \kappa_{\min(N_SM_R, N_SM_T)}\right) \in \mathbb{R}^{N_SM_R \times N_SM_T}$  is a diagonal matrix whose value on the diagonal is the singular value of **X**, and  $\kappa_1 \ge \kappa_2 \ge \cdots \ge \kappa_{\min(N_SM_R, N_SM_T)} \ge 0$ . The signal is expressed in a more compact form as:

$$\mathbf{X} = (\mathbf{W}_{R}\mathbf{B}\left(\boldsymbol{\phi}\right)) \operatorname{diag}\left(\mathbf{r}\right) \left(\mathbf{F}_{T}^{H}\mathbf{A}\left(\boldsymbol{\theta}\right)\right)^{H} + \mathbf{N}.$$
(27)

Suppose  $l = 1, 2, \dots, K$  the larger the *K* singular value and their homologous singular vectors are roughly determined by *K* paths, we can infer:

$$\mathbf{u}_{l} = \frac{\mathbf{W}_{R}\mathbf{b}\left(\phi_{l}\right)}{\left\|\mathbf{W}_{R}\mathbf{b}\left(\phi_{l}\right)\right\|_{2}}, \ \mathbf{v}_{l} = \frac{\mathbf{F}_{T}^{H}\mathbf{a}\left(\theta_{l}\right)}{\left\|\mathbf{F}_{T}^{H}\mathbf{a}\left(\theta_{l}\right)\right\|_{2}}$$
(28)

where  $\mathbf{u}_l$  and  $\mathbf{v}_l$  are the *l*<sup>th</sup> column of **U** and **V**, respectively; the singular value can be written as:

$$\kappa_{l} = |r_{l}| \left\| \mathbf{W}_{R} \mathbf{b} \left( \phi_{l} \right) \right\|_{2} \left\| \mathbf{F}_{T}^{H} \mathbf{a} \left( \theta_{l} \right) \right\|_{2}.$$
<sup>(29)</sup>

Then, the spatial angles are normalized by  $(\theta_l, \phi_l) \in \left\{\frac{(i-1)}{N_{BS}}, i = 1, 2, \dots, N_{BS}\right\}$  and assumed to lie in the quantized points; we have:

$$\mathbf{u}_{l} \approx \mathbf{W}_{R} \mathbf{E}_{R} \mathbf{b} \left( \phi_{l} \right), \ \mathbf{v}_{l} \approx \mathbf{F}_{T}^{H} \mathbf{E}_{T} \mathbf{a} \left( \theta_{l} \right)$$
(30)

where  $\mathbf{E}_R \in \mathbb{C}^{N_{BS} \times N_{BS}}$  and  $\mathbf{E}_T \in \mathbb{C}^{N_{BS} \times N_{BS}}$  are the discrete Fourier transform matrices; we can further obtain the coarse estimates of angle, which can be formulated as:

$$\phi_{l}^{(1)} = \frac{\arg\max_{q=1}^{N_{BS}} \mathbf{d}_{R}(q) - 1}{N_{BS}}$$

$$\theta_{l}^{(1)} = \frac{\arg\max_{q=1}^{N_{BS}} \mathbf{d}_{T}(q) - 1}{N_{BS}}$$
(31)

where:

$$\mathbf{d}_{R} = (\mathbf{W}_{R}\mathbf{E}_{R})^{H}\mathbf{W}_{R}\mathbf{E}_{R}\mathbf{b} (\phi_{l})$$
  
$$\mathbf{d}_{T} = \left(\mathbf{F}_{T}^{H}\mathbf{E}_{T}\right)^{H}\mathbf{F}_{T}^{H}\mathbf{E}_{T}\mathbf{a} (\theta_{l}).$$
(32)

At this point, we get the initial candidates for the spatial angle. Furthermore, due to the uncertainty of the number of paths of the estimated channel at the beginning, we set  $L_{init} \ge K$ ,  $l = 1, 2, \dots, L_{init}$  to ensure the accuracy of our estimated path, and some incorrect paths will be cut off in the proposed algorithm. The iterative search scheme of angles is explained in the next subsection.

### 3.3. Step-Length Optimization

In the previous section, it can be seen that the objective function  $Z^{(t)}(\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi})$  in (19) is the weighted sum of two parts, where  $\mu$  is the regularization parameter [39] that affects the tradeoff to some extent. In the proposed algorithm,  $\mu$  is simplified as:

$$\mu = \min\left[\left(\omega^{(t)}\right)^{-1} e, \ \mu_{\max}\right] \tag{33}$$

where *e* is a constant scale factor, and the squared residual at the *t*<sup>th</sup> iteration  $\omega^{(t)}$  is expressed as:

$$\omega^{(t)} = \left\| \mathbf{X} - \mathbf{W}_R \mathbf{B} \left( \boldsymbol{\phi}^{(t)} \right) diag \left( \mathbf{r}_{pe}^{(t)} \right) \mathbf{A}^H \left( \boldsymbol{\theta}^{(t)} \right) \mathbf{F}_T \right\|_F^2.$$
(34)

The main task is to search for new estimates  $\boldsymbol{\theta}^{(t+1)}$  and  $\boldsymbol{\phi}^{(t+1)}$  from the neighborhood of  $\boldsymbol{\theta}^{(t)}$  and  $\boldsymbol{\phi}^{(t)}$  at the *t*<sup>th</sup> iteration, so that the objective function *Z*<sup>(t)</sup> becomes smaller and eventually stabilizes. Considering the characteristics of the fast convergence speed of the Newton descent method and

the high accuracy of the steepest descent method, we try to suggest combining the two for better performance; this searching can be accomplished by:

$$\boldsymbol{\phi}^{(t+1)} = \boldsymbol{\phi}^{(t)} - \rho^{(t)} \cdot \mathbf{d}^{(t)}$$
  
$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \rho^{(t)} \cdot \mathbf{d}^{(t)}$$
(35)

where  $\rho^{(t)}$  is the step-length at the *t*<sup>th</sup> iteration and  $\mathbf{d}^{(t)}$  is the corresponding descent direction. To balance the high computational complexity of the Newton iterative algorithm, we consider using the Newton direction as the descending direction to achieve fast convergence in the first iteration, that is denote  $\mathbf{d}^{(t)} = \mathbf{d}_N^{(t)}$  when t = 1; the expression  $\mathbf{d}_N^{(t)}$  is:

$$\mathbf{d}_{N}^{(t)} = \nabla^{2} Z_{pe}^{(t)}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)})^{-1} \nabla Z_{pe}^{(t)}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)}).$$
(36)

After that, we choose the gradient direction to achieve high accuracy and set  $\mathbf{d}^{(t)} = \mathbf{d}^{(t)}_G$ ; the expression  $\mathbf{d}^{(t)}_G$  is:

$$\mathbf{d}_{G}^{(t)} = \nabla Z_{pe}^{(t)}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)})$$
(37)

it can be seen that the search direction involves the gradient calculation of the iterative function  $Z_{pe}$ . Taking the normalized angle  $\phi_k$  as an example, the calculation method is explained as follows.

Define 
$$\mathbf{J} = \sum_{i=1}^{N_{\mathrm{S}}M_{T}} \mathbf{Q}_{i}^{H} \mathbf{x}_{i}, \mathbf{D} = \mu^{-1}\mathbf{G} + \sum_{i=1}^{N_{\mathrm{S}}M_{T}} \mathbf{Q}_{i}^{H} \mathbf{Q}_{i}$$
, then  $Z_{pe}$  can be expressed as:  

$$Z_{pe} = \sum_{i=1}^{N_{\mathrm{S}}M_{T}} \mathbf{x}_{i}^{H} \mathbf{x}_{i} - \mathbf{J}^{H} \mathbf{D}^{-1} \mathbf{J}.$$
(38)

Take the partial derivative with respect to  $\phi_k$ , we can obtain:

$$\frac{\partial Z_{pe}}{\partial \phi_k} = -\frac{\partial \mathbf{J}^H}{\partial \phi_k} \mathbf{D}^{-1} \mathbf{J} - \mathbf{J}^H \frac{\partial \mathbf{D}^{-1}}{\partial \phi_k} \mathbf{J} - \mathbf{J}^H \mathbf{D}^{-1} \frac{\partial \mathbf{J}}{\partial \phi_k}$$
(39)

and  $\frac{\partial \mathbf{D}^{-1}}{\partial \phi_k} = -\mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \phi_k} \mathbf{D}^{-1}$ ; we have:

$$\frac{\partial Z_{pe}}{\partial \phi_k} = -\frac{\partial \mathbf{J}^H}{\partial \phi_k} \mathbf{D}^{-1} \mathbf{J} + \mathbf{J}^H \mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \phi_k} \mathbf{D}^{-1} \mathbf{J} - \mathbf{J}^H \mathbf{D}^{-1} \frac{\partial \mathbf{J}}{\partial \phi_k}$$
(40)

where:

$$\frac{\partial \mathbf{J}}{\partial \phi_k} = \sum_{i=1}^{N_S M_T} \frac{\partial \mathbf{Q}_i^H}{\partial \phi_k} \mathbf{x}_i$$
(41)

$$\frac{\partial \mathbf{D}}{\partial \phi_k} = \sum_{i=1}^{N_S M_T} \left( \frac{\partial \mathbf{Q}_i^H}{\partial \phi_k} \mathbf{Q}_i + \mathbf{Q}_i^H \frac{\partial \mathbf{Q}_i}{\partial \phi_k} \right)$$
(42)

$$\frac{\partial \mathbf{Q}_{i}}{\partial \phi_{k}} = \frac{\partial \mathbf{W}_{R} \mathbf{B}\left(\phi\right) diag\left(\mathbf{A}^{H}\left(\theta\right) \mathbf{f}_{\mathbf{T}}\right)}{\partial \phi_{k}} = \left[0 \cdots 0 \mathbf{W}_{R} \frac{\partial \mathbf{b}\left(\phi_{k}\right)}{\partial \phi_{k}} \mathbf{a}^{H}\left(\theta_{k}\right) \mathbf{f}_{\mathbf{T}} 0 \cdots 0\right].$$
(43)

Since there is only a certain transposition relationship between the gradient and the partial derivative, we have the gradient of  $Z_{pe}$ , and we can further obtain the second derivative:

$$\frac{\partial^{2} Z_{pe}}{\partial \phi_{k}^{2}} = -\frac{\partial^{2} \mathbf{J}^{H}}{\partial \phi_{k}^{2}} \mathbf{D}^{-1} \mathbf{J} + 2 \left( \frac{\partial \mathbf{J}^{H}}{\partial \phi_{k}} \mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \phi_{k}} \mathbf{D}^{-1} \mathbf{J} \right) - 2 \left( \frac{\partial \mathbf{J}^{H}}{\partial \phi_{k}} \mathbf{D}^{-1} \frac{\partial \mathbf{J}}{\partial \phi_{k}} \right) 
- 2 \left( \mathbf{J}^{H} \mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \phi_{k}} \mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \phi_{k}} \mathbf{D}^{-1} \mathbf{J} \right) + \mathbf{J}^{H} \mathbf{D}^{-1} \frac{\partial^{2} \mathbf{D}}{\partial \phi_{k}^{2}} \mathbf{D}^{-1} \mathbf{J} 
+ 2 \left( \mathbf{J}^{H} \mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \phi_{k}} \mathbf{D}^{-1} \frac{\partial \mathbf{J}}{\partial \phi_{k}} \right) - \mathbf{J}^{H} \mathbf{D}^{-1} \frac{\partial^{2} \mathbf{J}}{\partial \phi_{k}^{2}}$$
(44)

where:

$$\frac{\partial^2 \mathbf{J}}{\partial \phi_k^2} = \sum_{i=1}^{N_S M_T} \frac{\partial^2 \mathbf{Q}_i^H}{\partial \phi_k^2} \mathbf{x}_i$$
(45)

$$\frac{\partial^2 \mathbf{D}}{\partial \phi_k^2} = \sum_{i=1}^{N_S M_T} \left( \frac{\partial^2 \mathbf{Q}_i^H}{\partial \phi_k^2} \mathbf{Q}_i + 2 \frac{\partial \mathbf{Q}_i^H}{\partial \phi_k} \frac{\partial \mathbf{Q}_i}{\partial \phi_k} + \mathbf{Q}_i^H \frac{\partial^2 \mathbf{Q}_i}{\partial \phi_k^2} \right)$$
(46)

$$\frac{\partial^2 \mathbf{Q}_i}{\partial \phi_k^2} = \left[ 0 \cdots 0 \, \mathbf{W}_R \frac{\partial^2 \mathbf{b} \left(\phi_k\right)}{\partial \phi_k^2} \mathbf{a}^H \left(\theta_k\right) \mathbf{f_T} \, 0 \cdots 0 \right]. \tag{47}$$

Next, we think about whether we can optimize the iteration step-length in (35) to improve the speed of iteration. Considering that if the optimal step-length of  $\boldsymbol{\theta}$  and  $\boldsymbol{\phi}$  is directly solved, the algorithm complexity is very high, and the channel is only related to the path gain **r** and the spatial angles  $\boldsymbol{\theta}$  and  $\boldsymbol{\phi}$ , so when the gain has been optimized, we choose to use the optimal step-length of the channel to approximate the optimal step-length of the angles. The optimization method is as follows.

Denote  $\mathbf{x} = vec(\mathbf{X})$ ,  $\mathbf{\Omega} = \mathbf{F}_T^T \otimes \mathbf{W}_{\mathbf{R}}$ , and given the spatial angles  $\boldsymbol{\theta}^{(t)}$  and  $\boldsymbol{\phi}^{(t)}$  at the *t*<sup>th</sup> iteration, we have:

$$\mathbf{h}^{(t)} = vec\left(\mathbf{B}\left(\boldsymbol{\phi}^{(t)}\right) diag\left(\mathbf{r}_{pe}^{(t)}\right) \mathbf{A}^{H}\left(\boldsymbol{\theta}^{(t)}\right)\right). \tag{48}$$

Then, the cost function transformed from (34) is:

$$O^{(t)} = vec\left(\omega^{(t)}\right) = \left\|\mathbf{x} - \mathbf{\Omega}\mathbf{h}^{(t)}\right\|^{2}.$$
(49)

Therefore, such an optimization can be reformulated as:

$$\rho^{(t)} = \underset{\rho^{(t)}}{\arg\min} O\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right)$$
(50)

The key to this problem is how to get the suitable  $\rho^{(t)}$ ; define  $\lambda\left(\rho^{(t)}\right) = O\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right)$ , and take the derivative with respect to  $\lambda\left(\rho^{(t)}\right)$ ; we can get:

$$\frac{\partial \lambda \left( \boldsymbol{\rho}^{(t)} \right)}{\partial \boldsymbol{\rho}^{(t)}} = \nabla O \left( \mathbf{h}^{(t)} + \boldsymbol{\rho}^{(t)} \mathbf{d}^{(t)} \right)^{H} \mathbf{d}^{(t)}$$
(51)

where:

$$\nabla O\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right) = \nabla \left\|\mathbf{x} - \mathbf{\Omega}\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right)\right\|^{2}$$
  
=  $\nabla \left[\mathbf{x} - \mathbf{\Omega}\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right)\right]^{H} \left[\mathbf{x} - \mathbf{\Omega}\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right)\right]$  (52)  
=  $\mathbf{\Omega}^{H} \left[\mathbf{\Omega}\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right) - \mathbf{x}\right].$ 

Electronics 2020, 9, 301

Similarly, we have:

$$\nabla O\left(\mathbf{h}^{(t)}\right) = \mathbf{\Omega}^{H}\left(\mathbf{\Omega}\mathbf{h}^{(t)} - \mathbf{x}\right)$$
(53)

(52) can be simplified as:

$$\nabla O\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right) = \mathbf{\Omega}^{H}\mathbf{\Omega}\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right) - \mathbf{\Omega}^{H}\mathbf{x}$$
  
=  $\mathbf{\Omega}^{H}\mathbf{\Omega}\left(\mathbf{h}^{(t)} + \rho^{(t)}\mathbf{d}^{(t)}\right) + \nabla O\left(\mathbf{h}^{(t)}\right) - \mathbf{\Omega}^{H}\mathbf{\Omega}\mathbf{h}^{(t)}$   
=  $\rho^{(t)}\mathbf{\Omega}^{H}\mathbf{\Omega}\mathbf{d}^{(t)} + \nabla O\left(\mathbf{h}^{(t)}\right).$  (54)

Now, we can obtain:

$$\frac{\partial \lambda \left( \boldsymbol{\rho}^{(t)} \right)}{\partial \boldsymbol{\rho}^{(t)}} = \nabla O \left( \mathbf{h}^{(t)} \right)^{H} \mathbf{d}^{(t)} + \boldsymbol{\rho}^{(t)} \mathbf{d}^{(t)^{H}} \boldsymbol{\Omega}^{H} \boldsymbol{\Omega} \mathbf{d}^{(t)}.$$
(55)

Considering the complexity of the algorithm, the descent direction here is the gradient direction. Set  $\frac{\partial \lambda(\rho^{(t)})}{\partial \rho^{(t)}} = 0$ ; we can finally get:

$$\rho^{(t)} = -\frac{\nabla O\left(\mathbf{h}^{(t)}\right)^{H} \nabla O\left(\mathbf{h}^{(t)}\right)}{\nabla O\left(\mathbf{h}^{(t)}\right)^{H} \mathbf{\Omega}^{H} \mathbf{\Omega} \nabla O\left(\mathbf{h}^{(t)}\right)}.$$
(56)

From this, we get the approximate optimal step-length for the angle update. During the iterative searching, the spatial angle and path gain estimates can converge quickly, until the latest estimates are almost the same as the previous ones. The proposed channel estimation scheme is shown in Algorithm 1. We can achieve the purpose of accurate and fast channel estimation, which will be verified in subsequent simulations.

Since the Newton direction is used only once, the computational complexity of each iteration mainly lies in the calculation of the gradient direction and the optimization of the step-length. The complexity to compute the gradient direction is  $O(N_s^2 M_R M_T (N_{BS} + K) K^2)$ , and the complexity of the step calculation is  $O(K^2 N_{BS})$ . As a result, we can conclude that the proposed algorithm has the complexity  $O(N_s^2 M_R M_T (N_{BS} + K) K^2 + K^2 N_{BS})$ .

In the above algorithm, we successfully estimated the space angle  $\theta$ ,  $\phi$  and the combined gain **r**, where  $r_k = \alpha_k \beta_k$ . Next, the solution to the problem of how to separate the downlink gain  $\alpha_k$  from the combined gain is revealed.

Specifically, we adopt the unit symbol pilot  $s_k = 1$  sent by the UEs to the BS via the uplink channel, and the signal  $\mathbf{Y} \in \mathbb{C}^{N_S M_R \times 1}$  received by the BS is written as:

$$\mathbf{Y} = \mathbf{W}_{R} \left[ \mathbf{h}_{U}^{1} s_{1} + \mathbf{h}_{U}^{2} s_{2} + \dots + \mathbf{h}_{U}^{k} s_{k} \right]$$
  
=  $\mathbf{W}_{R} \left[ \beta_{1} \mathbf{b} \left( \phi_{1} \right) + \beta_{2} \mathbf{b} \left( \phi_{2} \right) + \dots + \beta_{k} \mathbf{b} \left( \phi_{k} \right) \right]$   
=  $\mathbf{T} \left[ \beta_{1}, \beta_{2}, \dots, \beta_{k} \right]^{T}$  (57)

where  $\mathbf{T} = \mathbf{W}_R [\mathbf{b}(\phi_1), \mathbf{b}(\phi_2), \cdots, \mathbf{b}(\phi_k)] \in \mathbb{C}^{N_S M_R \times K}$ ; when  $rank(\mathbf{T}) \ge K$ , we have  $\boldsymbol{\beta} = \mathbf{T}^{\dagger} \mathbf{Y}$ , where  $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_k]^T$ , we obtain uplink gain  $\beta_k$ , and the downlink gain  $\alpha_k$  can be separated by utilizing  $\alpha_k = \frac{r_k}{\beta_k}$ . Since  $\mathbf{W}_R$  can be easily designed as a matrix whose columns are full rank, the above conditions can be achieved.

10 of 18

**Input:** Received signals with noise **X**, orthogonal pilot signals  $\mathbf{S}_{m_T,m_R}$ , coding matrix  $\mathbf{W}_R$  and  $\mathbf{F}_T$ , the number of detection paths  $L_{init}$ , threshold  $\Gamma$  and  $\varepsilon$ .

**Output:** Estimated angles and gains of all paths.

First stage • Perform  $[\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}] = SVD(\mathbf{X});$ • Compute  $\mathbf{d}_{R} = (\mathbf{W}_{R}\mathbf{E}_{R})^{H}\mathbf{W}_{R}\mathbf{E}_{R}\mathbf{b}(\phi_{l}), \mathbf{d}_{T} = (\mathbf{F}_{T}^{H}\mathbf{E}_{T})^{H}\mathbf{F}_{T}^{H}\mathbf{E}_{T}\mathbf{a}(\theta_{l});$ • Initialize  $\phi_{l}^{(1)} = \frac{\arg\max_{q=1}^{N_{BS}}\mathbf{d}_{R}(q)-1}{N_{BS}}, \theta_{l}^{(1)} = \frac{\arg\max_{q=1}^{N_{BS}}\mathbf{d}_{T}(q)-1}{N_{BS}};$ • Estimate gain  $\mathbf{r}_{pe}^{(1)}$  by (25), and trim path if  $\mathbf{r}_{pe}^{(1)} < \Gamma$ Second stage 1. Repeat 2. Update  $\mu = \min\left[\left(\omega^{(t)}\right)^{-1}e, \mu_{\max}\right];$ 3. Build the function  $Z_{pe}^{(t)}(\boldsymbol{\theta}, \boldsymbol{\phi})$  by (26); 4. If t = 1, then 4. If t = 1, then 5.  $\mathbf{d}^{(t)} = \mathbf{d}_N^{(t)}, \mathbf{d}_N^{(t)} = \nabla^2 Z_{pe}^{(t)}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)})^{-1} \nabla Z_{pe}^{(t)}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)});$ 6. Else 7.  $\mathbf{d}^{(t)} = \mathbf{d}_{G}^{(t)}, \mathbf{d}_{G}^{(t)} = \nabla Z_{pe}^{(t)}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)});$ 8. End if 9. Step optimization calculation: 9.1  $\mathbf{x} = vec(\mathbf{X}), \mathbf{h}^{(t)} = vec(\mathbf{H}^{(t)});$ 9.2 Compute  $\nabla O\left(\mathbf{h}^{(t)}\right) = \mathbf{\hat{\Omega}}^{H}\left(\mathbf{\Omega}\mathbf{h}^{(t)} - \mathbf{x}\right);$ 9.3 Compute  $\rho^{(t)} = -\frac{\nabla O(\mathbf{h}^{(t)})^H \nabla O(\mathbf{h}^{(t)})}{\nabla O(\mathbf{h}^{(t)})^H \Omega^H \Omega \nabla O(\mathbf{h}^{(t)})}.$ 10.  $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \rho^{(t)} \cdot \mathbf{d}^{(t)}, \boldsymbol{\phi}^{(t+1)} = \boldsymbol{\phi}^{(t)} - \rho^{(t)} \cdot \mathbf{d}^{(t)};$ 11. Estimate the path gain  $\mathbf{r}_{pe}^{(t+1)}$ , and streamline path if  $\mathbf{r}_{pe}^{(t+1)} < \Gamma$ ; 12. Until  $K^{(t)} = K^{(t+1)}$ , and  $\|\mathbf{r}^{(t+1)} - \mathbf{r}^{(t)}\|_2 < \varepsilon$ .

### 4. Simulation Results

In this section, we demonstrate the simulation comparisons under different parameter scenarios, which can prove the effectiveness of the proposed step-length optimization-based joint iterative scheme. In particular, simulation results are provided to verify that the proposed algorithm can outperform the previous methods in certain aspects, as in [26,27]. Some default parameters are set to d = 0.5,  $N_S = 4$ , and L = 24. Moreover, the number of antennas, UEs, and time slots is set to  $N_{BS} = 64$ , K = 3, and  $M_R = M_T = 12$ , respectively, which may change in the form of variables in subsequent simulations, and we set the variable  $M_P$  to represent the time slot for better presentation. As in [2,40], the signal-to-noise ratio (SNR) is defined as  $\sigma_s^2 / \sigma_n^2$ , where  $\sigma_s^2$  is the signal power. *I* denotes the number of Monte Carlo runs, and the normalized mean squared error (NMSE) is defined as:

$$NMSE = \frac{1}{I} \left( \sum_{i=1}^{I} \frac{\left\| \hat{\mathbf{H}}_i - \mathbf{H}_i \right\|_F^2}{\left\| \mathbf{H}_i \right\|_F^2} \right)$$
(58)

Figure 2 shows the convergence performance of the optimized scheme compared with the conventional super-resolution (SR) algorithm, which is reflected by the relationship between NMSE and the number of iterations for three SNR values. It is obvious that the proposed scheme can achieve convergence with a small number of iterations for each SNR value. For instance, the proposed algorithm completes convergence in about 20 iterations at an SNR of 20 dB, which also fully proves the acceleration brought by optimization.



Figure 2. NMSE versus the number of iterations. SR, super-resolution.

Figure 3 compares the NMSE performance against the SNR. We assume the system design parameters  $N_{BS} = 32$  and  $M_R = M_T = M_P$ , as shown in Figure 3, and the accuracy of the proposed solution is similar to that of the SR algorithm in different time slots, which proves that the proposed optimization does not sacrifice performance indicators. We can see that there is a certain proportional relationship between the time slot and the training pilot, so the NMSE performance becomes better as the number of time slots increases. In addition, the NMSE of the proposed and SR schemes is also shown in Table 1 for further comparison. It is clearly observed that accurate channel estimation can be achieved.



Figure 3. NMSE performance comparison of the proposed and SR algorithms.

| (SNR)<br>(dB)           | 0         | 5         | 10        | 15        | 20        | 25        | 30        |
|-------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $SR, M_P = 6$           | 4.4983e-3 | 1.2481e-3 | 3.4718e-4 | 1.0425e-4 | 3.3272e-5 | 1.1428e-5 | 5.9694e-6 |
| Proposed,<br>$M_P = 6$  | 4.2355e-3 | 1.1838e-3 | 3.2268e-4 | 1.0394e-4 | 3.5274e-5 | 1.1674e-5 | 5.7921e-6 |
| $SR, M_P = 8$           | 2.0730e-3 | 6.0159e-4 | 1.9159e-4 | 5.9361e-5 | 1.8088e-5 | 5.8977e-6 | 2.0542e-6 |
| Proposed,<br>$M_P = 8$  | 2.1037e-3 | 6.0318e-4 | 1.8060e-4 | 5.7340e-5 | 1.7180e-5 | 5.5994e-6 | 2.1856e-6 |
| $SR, M_P = 10$          | 1.3365e-3 | 3.9811e-4 | 1.2478e-4 | 4.1041e-5 | 1.2613e-5 | 4.1339e-6 | 1.3255e-6 |
| Proposed,<br>$M_P = 10$ | 1.3688e-3 | 3.7503e-4 | 1.2548e-4 | 3.9852e-5 | 1.2220e-5 | 3.9532e-6 | 1.3329e-6 |
| $SR, M_P = 12$          | 8.0790e-4 | 2.4650e-4 | 8.0040e-5 | 2.7415e-5 | 8.0607e-6 | 2.6785e-6 | 8.4921e-7 |
| Proposed,<br>$M_P = 12$ | 8.1252e-4 | 2.2682e-4 | 7.9436e-5 | 2.5461e-5 | 8.3286e-6 | 2.5191e-6 | 7.7909e-7 |

Table 1. NMSE of the proposed scheme and SR algorithms in different time slots.

Figure 4 depicts the iteration time of the proposed algorithm on the basis of the hypothetical scenario in Figure 3, which can more intuitively reflect that the speed of the proposed scheme is significantly improved and the computational complexity is reduced. For example, when the SNR is 20 and the time slot is 10, the time taken by the proposed algorithm is 2.0534 s and the SR is 5.0885 s; it can be fully seen that our method achieves fast estimation. We can also see that as the SNR and time slot increase, the time consumed increase, and that is because the accuracy is also improved.



Figure 4. Comparison of the proposed and SR algorithms in terms of processing time.

Figure 5 displays the NMSE performance comparison with BCS in the case of different numbers of UEs. Since it optimizes all the spatialization angles simultaneously and eliminates some false paths, the proposed scheme can obtain better NMSE performance, and the processing time of the BCS method is significantly higher than the proposed scheme under the same conditions. In addition, we can see that when the number of users increases, the channel estimation accuracy of the proposed algorithm decreases; but the decrease is small, and the proposed algorithm can still effectively estimate the channel.



Figure 5. NMSE performance comparison of the proposed and BCS algorithms.

Figure 6 compares the angular mean squared error (RMSE) performance under various UEs and antennas. The RMSE is given by:

$$RMSE = \frac{1}{K} \sum_{k=1}^{K} \sqrt{\frac{1}{I} \sum_{i=1}^{I} \left[ \left( \hat{\theta}_{k,i} - \theta_k \right)^2 + \left( \hat{\phi}_{k,i} - \phi_k \right)^2 \right]}$$
(59)

As SNR increases, we observe that the RMSE of SVD remains constant, and its computational complexity is lower; but our algorithm yields the best performance. This is because the SNR is related to the entire angle domain grids, and our SVD method finds the best matching grid to get coarse estimates. Similarly to Figure 5, the RMSE performance of channel estimation will decrease with the increase of users, and in contrast, the number of antennas is directly proportional to the estimated accuracy.



Figure 6. Cont.



**Figure 6.** RMSE performance comparison of the proposed and SVD algorithms: (**a**) RMSE versus the number of UEs; (**b**) RMSE versus the number of antennas.

Figure 7 shows the performance of the gain mean squared error (GMSE) of the proposed algorithm in different time slots, and GMSE is expressed as  $GMSE = \frac{1}{K} \sum_{k=1}^{K} \sqrt{\frac{1}{I} \sum_{i=1}^{I} (\hat{\alpha}_{k,j} - \alpha_k)^2}$ . Obviously, the recovery of the channel can gradually improve with the increase of time slots and SNR.



Figure 7. GMSE performance of the proposed algorithm with different time slots.

## 5. Conclusions

In this paper, we proposed a joint iterative channel estimation scheme based on step-length optimization for multi-user mmWave massive MIMO systems. The proposed scheme first combined the advantages of Newton method and gradient descent method, and then gave a step-length optimization approach. Compared with the conventional SR algorithm, the proposed method greatly reduced the computational complexity without losing estimation accuracy, and the proposed algorithm

yielded smaller channel estimation error compared to existing BCS and SVD methods. In addition, the proposed method could effectively estimate the channel even with a certain number of UEs. The perspective of this work considered an extension to the mmWave time-varying channel for joint channel parameters estimation, which included AoDs/AoAs, path gains, and Doppler shifts.

**Author Contributions:** Conceptualization J.D.; methodology, J.D.; software, J.L. and H.L.; validation, J.L. and J.H.; writing, original draft preparation, J.D. and J.L.; writing, review and editing, J.H., Y.G., and H.L.; supervision, J.D. and Y.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the grant from the National Natural Science Foundation of China (Nos. 61601414 and 61702466), the National Key Research and Development Program of China (No. 2016YFB0502001), and the Fundamental Research Funds for the Central Universities (No. 2018CUCTJ082).

**Acknowledgments:** The authors would like to thank the anonymous reviewers and the Editor for their careful reviews and constructive suggestions to help us improve the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

## References

- 1. Gao, Z.; Hu, C.; Dai, L.; Wang, Z. Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels. *IEEE Commun. Lett.* **2016**, *20*, 1259–1262. [CrossRef]
- 2. Wang, M.; Gao, F.; Shlezinger, N.; Flanagan, M.F.; Eldar, Y.C. A block sparsity based estimator for mmWave massive MIMO channels with beam squint. *IEEE Trans. Signal Process.* **2019**, *68*, 49–64. [CrossRef]
- 3. Ding, T.; Zhao, Y.; Li, L. Overlapped Sub-Array Based Hybrid Precoding for mmWave Massive MIMO System with Lens Arrays. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2019, pp. 537–541.
- Xiao, M.; Mumtaz, S.; Huang, Y.; Dai, L.; Li, Y.; Matthaiou, M.; Karagiannidis, G.K.; Björnson, E.; Yang, K.; Chih-Lin, I.; et al. Millimeter wave communications for future mobile networks. *IEEE J. Sel. Areas Commun.* 2017, 35, 1909–1935. [CrossRef]
- Rappaport, T.S.; MacCartney, G.R.; Samimi, M.K.; Sun, S. Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design. *IEEE Trans. Commun.* 2015, *63*, 3029–3056. [CrossRef]
- 6. Marzetta, T.L. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 3590–3600. [CrossRef]
- Ghallab, R.; Shokair, M.; Elazm, A.A.; Sakr, A.; Saad, W.; Naguib, A. Performance Enhancement Using MIMO Electronic Relay in Massive MIMO Cellular Networks. *IET Netw.* 2019, *8*, 299–306. [CrossRef]
- 8. Hoydis, J.; Ten Brink, S.; Debbah, M. Massive MIMO in the UL/DL of cellular networks: How many antennas do we need? *IEEE J. Sel. Areas Commun.* **2013**, *31*, 160–171. [CrossRef]
- 9. Alibakhshikenari, M.; Virdee, B.S.; Shukla, P.; See, C.H.; Abd-Alhameed, R.A.; Falcone, F.; Quazzane, K.; Limiti, E. Isolation enhancement of densely packed array antennas with periodic MTM-photonic bandgap for SAR and MIMO systems. *IET Microwaves, Antennas & Propag.* **2019**. [CrossRef]
- Alibakhshikenari, M.; Virdee, B.S.; See, C.H.; Abd-Alhameed, R.A.; Falcone, F.; Limiti, E. Surface wave reduction in antenna arrays using metasurface inclusion for MIMO and SAR systems. *Radio Sci.* 2019, 54, 1067–1075. [CrossRef]
- 11. Zhang, Q.; Jin, S.; Wong, K.K.; Zhu, H.; Matthaiou, M. Power scaling of uplink massive MIMO systems with arbitrary-rank channel means. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 966–981. [CrossRef]
- 12. Chen, K.; Qi, C.; Li, G.Y. Two-Step Codeword Design for Millimeter Wave Massive MIMO Systems with Quantized Phase Shifters. *IEEE Trans. Signal Process.* **2019**, *68*, 170–180. [CrossRef]
- 13. Zhang, Y.; Du, J.; Chen, Y.; Han, M.; Li, X. Optimal Hybrid Beamforming Design for Millimeter-Wave Massive Multi-User MIMO Relay Systems. *IEEE Access* **2019**, *7*, 157212–157225. [CrossRef]
- 14. González-Coma, J.P.; Rodriguez-Fernandez, J.; González-Prelcic, N.; Castedo, L.; Heath, R.W. Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 353–367. [CrossRef]
- 15. Chen, L.; Yuan, X. Blind Multiuser Detection in Massive MIMO Channels with Clustered Sparsity. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 1052–1055. [CrossRef]

- 16. Zhang, J.; Yuan, X.; Zhang, Y.J.A. Blind signal detection in massive MIMO: Exploiting the channel sparsity. *IEEE Trans. Commun.* **2017**, *66*, 700–712. [CrossRef]
- 17. Swindlehurst, A.L.; Ayanoglu, E.; Heydari, P.; Capolino, F. Millimeter-wave massive MIMO: The next wireless revolution? *IEEE Commun. Mag.* 2014, *52*, 56–62. [CrossRef]
- Fan, D.; Gao, F.; Liu, Y.; Deng, Y.; Wang, G.; Zhong, Z.; Nallanathan, A. Angle Domain Channel Estimation in Hybrid Millimeter Wave Massive MIMO Systems. *IEEE Trans. Wirel. Commun.* 2018, 17, 8165–8179. [CrossRef]
- 19. Shao, W.; Zhang, S.; Zhang, X.; Ma, J.; Zhao, N.; Leung, V.C. Massive MIMO Channel Estimation Over the mmWave Systems Through Parameters Learning. *IEEE Commun. Lett.* **2019**, *23*, 672–675. [CrossRef]
- 20. Alkhateeb, A.; El Ayach, O.; Leus, G.; Heath, R.W. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 831–846. [CrossRef]
- 21. Xie, H.; Gao, F.; Jin, S. An overview of low-rank channel estimation for massive MIMO systems. *IEEE Access* **2016**, *4*, 7313–7321. [CrossRef]
- 22. Bellili, F.; Sohrabi, F.; Yu, W. Generalized Approximate Message Passing for Massive MIMO mmWave Channel Estimation With Laplacian Prior. *IEEE Trans. Commun.* **2019**, *67*, 3205–3219. [CrossRef]
- Lee, B.; Choi, J.; Seol, J.y.; Love, D.J.; Shim, B. Antenna grouping based feedback reduction for FDD-based massive MIMO systems. In Proceedings of the 2014 IEEE International Conference on Communications (ICC), Sydney, Australia, 4–14 June 2014; pp. 4477–4482.
- 24. Shen, W.; Dai, L.; Shim, B.; Mumtaz, S.; Wang, Z. Joint CSIT acquisition based on low-rank matrix completion for FDD massive MIMO systems. *IEEE Commun. Lett.* **2015**, *19*, 2178–2181. [CrossRef]
- 25. Fang, J.; Wang, F.; Shen, Y.; Li, H.; Blum, R.S. Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach. *IEEE Trans. Signal Process.* **2016**, *64*, 4649–4662. [CrossRef]
- 26. Hu, C.; Dai, L.; Mir, T.; Gao, Z.; Fang, J. Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding. *IEEE Trans. Veh. Technol.* **2018**, *67*, 8954–8958. [CrossRef]
- 27. Wu, Y.; Gu, Y.; Wang, Z. Channel estimation for mmWave MIMO with transmitter hardware impairments. *IEEE Commun. Lett.* **2017**, *22*, 320–323. [CrossRef]
- 28. Liao, A.; Gao, Z.; Wu, Y.; Wang, H.; Alouini, M.S. 2D unitary ESPRIT based super-resolution channel estimation for millimeter-wave massive MIMO with hybrid precoding. *IEEE Access* **2017**, *5*, 24747–24757. [CrossRef]
- 29. El Ayach, O.; Rajagopal, S.; Abu-Surra, S.; Pi, Z.; Heath, R.W. Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 1499–1513. [CrossRef]
- Xia, P.; Yong, S.K.; Oh, J.; Ngo, C. A practical SDMA protocol for 60 GHz millimeter wave communications. In Proceedings of the 2008 42nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 26–29 October 2008; pp. 2019–2023.
- 31. Tse, D.; Viswanath, P. *Fundamentals of Wireless Communication*; Cambridge University Press: Cambridge, UK, 2005.
- 32. Wang, M.; Gao, F.; Jin, S.; Lin, H. An overview of enhanced massive MIMO with array signal processing techniques. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 886–901. [CrossRef]
- 33. Qin, Q.; Gui, L.; Cheng, P.; Gong, B. Time-varying channel estimation for millimeter wave multiuser MIMO systems. *IEEE Trans. Veh. Technol.* **2018**, *67*, 9435–9448. [CrossRef]
- Mumtaz, S.; Rodriguez, J.; Dai, L. MmWave Massive MIMO: A Paradigm for 5G; Academic Press: Cambridge, MA, USA, 2016.
- 35. Shen, Y.; Fang, J.; Li, H. Exact reconstruction analysis of log-sum minimization for compressed sensing. *IEEE Signal Process. Lett.* **2013**, *20*, 1223–1226. [CrossRef]
- 36. Lange, K.; Hunter, D.R.; Yang, I. Optimization transfer using surrogate objective functions. *J. Comput. Graph. Stat.* **2000**, *9*, 1–20.
- 37. Tang, Y.; Vucetic, B.; Li, Y. An iterative singular vectors estimation scheme for beamforming transmission and detection in MIMO systems. *IEEE Commun. Lett.* **2005**, *9*, 505–507. [CrossRef]
- Jiang, T.; Xie, F.; Yuan, S.; Yao, S.; Wu, K.; Wang, X. Implementation of Matrix SVD Decomposition Module for Subspace Channel Estimation. In Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 15–17 March 2019; pp. 1096–1102.

- Bauer, F.; Lukas, M.A. Comparingparameter choice methods for regularization of ill-posed problems. *Math. Comput. Simul.* 2011, 81, 1795–1841. [CrossRef]
- 40. Liu, K.; Da Costa, J.P.C.; So, H.C.; De Almeida, A.L. Semi-blind receivers for joint symbol and channel estimation in space-time-frequency MIMO-OFDM systems. *IEEE Trans. Signal Process.* **2013**, *61*, 5444–5457. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).