Channel Estimation for Intelligent Reflecting Surface Assisted mmWave Communications

Dong Chen (✉ cd906253519@163.com)

Nanjing University of Posts and Telecommunications <https://orcid.org/0000-0003-3640-2450>

Youhua Fu

Nanjing University of Posts and Telecommunications

Chen Liu

Nanjing University of Posts and Telecommunications

Hong Wang

Nanjing University of Posts and Telecommunications

Research

Keywords: Intelligent reflecting surface, Millimeter wave, Regular alternating least squares, Channel estimation, Convex optimization

Posted Date: September 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-853256/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Channel Estimation for Intelligent Reflecting Surface Assisted mmWave Communications

Dong Chen*, Youhua Fu, Chen Liu, and Hong Wang

*Correspondence:cd906253519@163.com

Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract

Intelligent reflecting surface (IRS) consists of a large number of low-cost passive reflective elements, which can assist millimeter wave communications to solve the problems of weak penetration and short propagation distance. However, it is challenging to obtain channel state information (CSI) in IRS-aided millimeter wave communication systems. To solve this challenge, this paper proposes a regular alternating least squares (RALS) algorithm based on the canonical/parallel factor (CP) decomposition. Compared with the traditional alternate least squares (ALS) algorithm, the proposed RALS algorithm has better convergence performance, thus solving the problem of divergence or slow convergence of the conventional ALS algorithm. Besides, in order to improve the accuracy of the channel estimation, the convex optimization theory is invoked to devise the regularization parameters, and a regularization parameter selection scheme is developed to ensure that the proposed algorithm obtains the optimal solution. The simulation results verify the theoretical analysis and prove the superiority of the proposed RALS algorithm in terms of estimation error performance.

Keywords: Intelligent reflecting surface, Millimeter wave, Regular alternating least squares, Channel estimation, Convex optimization

1 Introduction

Massive multiple-input-multiple-output (MIMO) technology is used in 5G communication systems to compensate for millimeter wave (mmWave) propagation loss [1]. In order to further improve the system performance, intelligent reflecting surface (IRS) is perceived as a promising technique for mmWave communications. IRS is a plane composed of a large number of passive reflective elements. On the one hand, it can adjust the amplitude and phase of incident signal intelligently to achieve the effect of reconfiguring the propagation environment [2]-[3]. On the other hand, it can improve the spectrum efficiency and energy efficiency of wireless communications in a cost-effective way [4]-[6]. These benefits of IRS can be combined with some communication techniques, e.g., orthogonal frequency division multiplexing (OFDM) [7]-[8], non-orthogonal multiple access (NOMA) [9]-[10], and so on. However, the introduction of IRS makes the channel more

complicated and brings certain difficulties to channel estimation. Specifically, due to the passive nature of IRS, it is quite difficult to estimate the channel from base station (BS) to IRS and the channel from IRS to mobile station (MS) separately.

In order to overcome the above difficulties, we can use separate estimation of the BS-IRS channel and the IRS-MS channel scheme or cascaded channel estimation scheme. In the following, we will discuss these two types of methods, respectively.

First, we discuss the existing work about estimating the BS-IRS and IRS-MS channels separately. In [11], the researchers proposed a new IRS hardware structure, in which a few active reflection elements were randomly deployed in the IRS. According to this structure, a channel estimation scheme based on compressed sensing (CS) and deep learning (DL) was proposed, in which the full channels between the IRS and the BS/MS were estimated from the available channel state information (CSI) sensed at the active elements. Although this scheme can easily estimate the BS-IRS channel and the IRS-MS channel separately, in the presence of the active elements, it will increase the cost and energy consumption. Besides, in [12], a tensor model was established by using its algebraic structure, so that the two channels can be estimated separately. Compared with the scheme in [11], this scheme reduces the cost and energy consumption at the expense of the channel estimation accuracy.

Second, we review the existing work about cascaded channel estimation schemes. In [7]-[8], in order to reduce training overhead and simplify the design of passive beamforming, a new elements grouping strategy is proposed, which only needs to estimate the cascaded channels associated with each subsurface; In [13], the channel estimation problem was transformed into a compressed sensing (CS) problem, and a new low-complexity algorithm based on CS was proposed; In [14], the investigators elaborated on the estimation of cascaded channels, and proposed a channel estimation algorithm based on sparse matrix decomposition and complete matrix. The above estimation schemes can easily estimate the cascaded channel, but they have the common limitation, i.e., the specific CSI of the BS-IRS channel or the IRS-MS channel can not be obtained.

In this paper, IRS is applied to the millimeter wave communication system. In order to obtain the specific SCI of the BS-IRS channel and the IRS-MS channel, this paper proposes a regular alternating least squares (RALS) algorithm. This algorithm first establishes the relationship between the received signal and the tensor decomposition to satisfy the CP decomposition model, then obtains two forms of the horizontally expanded CP decomposition. Finally, estimate the BS-IRS channel and the IRS-MS channel according to the expanded forms. In order to avoid the problem that the separate estimation will reduce the accuracy of the estimation, the algorithm adds regularization parameters, then uses convex optimization theory to analyze them to get the most suitable regularization parameters, finally obtains the optimal solution. The advantage of this algorithm is that it has a faster convergence rate than the traditional alternating least squares algorithm, avoids divergence, and also solves the problem of data instability caused by singular value decomposition when solving pseudo-inverses. The simulation results also verify the effectiveness of the proposed scheme. In summary, the contributions of this paper are listed as follows:

- (1) We propose a channel estimation scheme called RALS. By using CP decomposition, our proposed RALS algorithm can realize the separate estimation of the BS-IRS channel and the IRS-MS channel. Due to the addition of regularization parameters, it can speed up the convergence speed of the result and avoid divergence.
- (2) In order to improve the accuracy of estimation, we use convex optimization theory to analyze the regularization parameters to obtain the optimal parameters. Ensure that the RALS algorithm is a convex problem, and the optimal solution can be obtained.

2 System model

This paper considers a downlink IRS assisted millimeter wave communication systems. The IRS is composed of N_{IRS} passive reflective elements, in which the phase shifts and the amplitudes can be adjusted independently. Both the transmitter and the receiver of the system are equipped with multiple antennas. That is, the base station (BS) and the mobile station (MS) are equipped with M and N antennas respectively. In this case, the communication system has two propagation links, which are the BS directly to the MS and the BS to the MS through the IRS. There are three kinds of channels on these two links:

the channel from the BS directly to the MS $\mathbf{L}$, the channel from the BS to the IRS $\mathbf{H}_1$, and the channel from the IRS to the MS $\mathbf{H}_2$. When the direct path is blocked by the obstacles, just the path provided by IRS is used, which can help millimeter wave transmission. This article assumes that the direct path is blocked by the obstacles. The considered IRS-assisted millimeter wave communication system model is shown in Fig. 1.

Fig. 1 IRS assisted millimeter wave communication system

In the channel estimation phase of the considered IRS assisted millimeter wave communications, it is assumed that there are T consecutive time frames, and each time frame contains Q time slots. Without loss of generality, the following assumptions are made: the phase shift matrix of the IRS in each slot of the same time frame remains invariable, and only alters with the time frames. Moreover, the pilot symbol is repeated on the T different time frames.

Therefore, the received signal at the MS of the q -th ($q \in \{1, \dots, Q\}$) time slot in the t -th ($t \in \{1, \dots, T\}$) time frame can be written as

$$y_{t,q} = \mathbf{H}_2 \text{diag}(\mathbf{P}_t) \mathbf{H}_1 \mathbf{x}_q + \mathbf{n}_{t,q} \quad (1)$$

where $\mathbf{H}_1 \in \mathbb{C}^{N_{\text{IRS}} \times M}$ is the channel between the BS and the IRS, $\mathbf{H}_2 \in \mathbb{C}^{N \times N_{\text{IRS}}}$ is the channel between the IRS and MS.

$\text{diag}(\mathbf{P}_t) = \text{diag}([a_1 e^{j\theta_1}, a_2 e^{j\theta_2}, \dots, a_{N_{\text{IRS}}} e^{j\theta_{N_{\text{IRS}}}}]^T) \in \mathbb{C}^{N_{\text{IRS}} \times N_{\text{IRS}}}$ is the phase shift matrix of on the IRS, where a_n ($n \in \{1, \dots, N_{\text{IRS}}\}$) represents the amplitude and $\theta_n \in (0, 2\pi]$ represents the phase shift. In order to facilitate the implementation in practice, this paper only studies the phase shift changes, and all the amplitudes are set to be ones, i.e., $a_n = 1$ ($n \in \{1, \dots, N_{\text{IRS}}\}$). $\mathbf{x}_q \in \mathbb{C}^{M \times 1}$ is the pilot symbol vector transmitted at the q -th slot in time frame t . $\mathbf{n}_{t,q} \in \mathbb{C}^{N \times 1}$ is the additive white Gaussian noise (AWGN) vector.

Accordingly, the received signal in the t -th frame is expressed as

$$\mathbf{Y}_t = \mathbf{H}_2 \Phi_t(\mathbf{P}) \mathbf{H}_1 \mathbf{X} + \mathbf{N}_t \quad (2)$$

where $\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_T]^T \in \mathbb{C}^{T \times N_{IRS}}$, $\Phi_t(\mathbf{P}) = \text{diag}(\mathbf{P}_t)$ denotes a diagonal matrix with the t -th row of the IRS phase shift matrix $\mathbf{P}$ as its main diagonal, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_Q] \in \mathbb{C}^{M \times Q}$ is the pilot symbol matrix in the entire time frame t , and $\mathbf{N}_t = [\mathbf{n}_{t,1}, \dots, \mathbf{n}_{t,Q}] \in \mathbb{C}^{N \times Q}$ is the AWGN matrix.

From Eq. (2), we know that $\mathbf{Y}_t$ is the received signal in the t -th time frame, and thus the received signal in T time frames can be written as

$$\mathcal{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_T] \in \mathbb{C}^{N \times Q \times T} \quad (3)$$

Finally, we analyze the channel model. Due to the existence of scattering paths, the millimeter wave channel has rich geometric features, in which each scattering path should determine a single propagation path based on the geometric channel model.

Under this model, the channel between BS-IRS and the channel between IRS-MS can be respectively expressed as

$$\mathbf{H}_1 = \sqrt{\frac{MN_{IRS}}{L_1}} \sum_{l_1=1}^{L_1} \alpha_{l_1} a_{IRS}(\gamma_{l_1}, \phi_{l_1}) a_{BS}^H(\gamma_{l_1}', \phi_{l_1}') \quad (4)$$

$$\mathbf{H}_2 = \sqrt{\frac{NN_{IRS}}{L_2}} \sum_{l_2=1}^{L_2} \beta_{l_2} b_{MS}(\gamma_{l_2}, \phi_{l_2}) b_{IRS}^H(\gamma_{l_2}', \phi_{l_2}') \quad (5)$$

where α_{l_1} (β_{l_2}) is the path gain of the l_1 (l_2)-th path. L_1 and L_2 are the number of scattering paths on the channel $\mathbf{H}_1$ and the channel $\mathbf{H}_2$, respectively. Some previous papers assumed that the antennas follow a uniform linear array (ULA) structure for research simplicity, but this is not in reality situation. In this paper, the uniform planar array (UPA) structure is used. Then the response vectors of the antenna array in the channel can be respectively written as

$$a_{IRS}(\gamma_{l_1}, \phi_{l_1}) = \frac{1}{\sqrt{N_{IRS}}} \left[1, \dots, e^{j \frac{2\pi}{\lambda} d (m \sin(\gamma_{l_1}) \sin(\phi_{l_1}) + n \cos(\phi_{l_1}))}, \dots, e^{j \frac{2\pi}{\lambda} d ((\sqrt{N_{IRS}}-1) \sin(\gamma_{l_1}) \sin(\phi_{l_1}) + (\sqrt{N_{IRS}}-1) \cos(\phi_{l_1}))} \right]^T \quad (6)$$

$$a_{BS}(\gamma_{l_1}', \phi_{l_1}') = \frac{1}{\sqrt{M}} \left[1, \dots, e^{j \frac{2\pi}{\lambda} d (m \sin(\gamma_{l_1}') \sin(\phi_{l_1}') + n \cos(\phi_{l_1}'))}, \dots, e^{j \frac{2\pi}{\lambda} d ((\sqrt{M}-1) \sin(\gamma_{l_1}') \sin(\phi_{l_1}') + (\sqrt{M}-1) \cos(\phi_{l_1}'))} \right]^T \quad (7)$$

where λ is the millimeter wave wavelength. d is the spacing distance between the antenna arrays or that between the elements on the IRS with $d = \lambda / 2$. γ_{l_1}' and ϕ_{l_1}' are the IRS (BS) azimuth angle and elevation angle, respectively.

In order to make the structure more compact, we define three new variables, which are given as

$$\alpha = \sqrt{\frac{MN_{IRS}}{L_1}} [\alpha_1, \alpha_2, \dots, \alpha_{L_1}] \quad (8)$$

$$\mathbf{A}_{BS} = [a_{BS}(\gamma'_1, \phi'_1), a_{BS}(\gamma'_2, \phi'_2), \dots, a_{BS}(\gamma'_{L_1}, \phi'_{L_1})] \quad (9)$$

$$\mathbf{A}_{IRS} = [a_{IRS}(\gamma_1, \phi_1), a_{IRS}(\gamma_2, \phi_2), \dots, a_{IRS}(\gamma_{L_1}, \phi_{L_1})] \quad (10)$$

Therefore, according to Eq. (8), Eq. (9) and Eq. (10), Eq. (4) can be rewritten as

$$\mathbf{H}_1 = \mathbf{A}_{IRS} \text{diag}(\alpha) \mathbf{A}_{BS}^H \quad (11)$$

Similar to the channel $\mathbf{H}_1$, the expression of the channel $\mathbf{H}_2$ can be rewritten as

$$\mathbf{H}_2 = \mathbf{B}_{MS} \text{diag}(\beta) \mathbf{B}_{IRS}^H \quad (12)$$

where

$$\beta = \sqrt{\frac{NN_{IRS}}{L_2}} [\beta_1, \beta_2, \dots, \beta_{L_2}] \quad (13)$$

$$\mathbf{B}_{MS} = [b_{MS}(\gamma'_1, \phi'_1), b_{MS}(\gamma'_2, \phi'_2), \dots, b_{MS}(\gamma'_{L_2}, \phi'_{L_2})] \quad (14)$$

$$\mathbf{B}_{IRS} = [b_{IRS}(\gamma_1, \phi_1), b_{IRS}(\gamma_2, \phi_2), \dots, b_{IRS}(\gamma_{L_2}, \phi_{L_2})] \quad (15)$$

3 Proposed channel estimation algorithm

Currently, most of the previous papers focused on the design of the reflection matrix for millimeter wave communication system, in which it was assumed that either the CSI is fully known at the BS side or the MS side, or the channel information between the BS and the IRS is fully known. However, since the elements on the IRS are passive, it is challenging to obtain channel state information in practice. Therefore, this paper intends to estimate the channel $\mathbf{H}_1$ between BS-IRS and the channel $\mathbf{H}_2$ between IRS-MS at the same time based on the channel model developed in the last section.

It can be seen from Eq. (3) that the received signal is a third-order tensor, which makes the channel estimation analysis and calculation more complicated. To proceed, we need to establish the relationship between the received signal and the tensor decomposition.

Letting $\mathbf{W} = (\mathbf{H}_1 \mathbf{X})^H \in \mathbb{C}^{Q \times N_{IRS}}$, Eq. (2) can be rewritten as

$$\mathbf{Y}_t = \mathbf{H}_2 \Phi_t(\mathbf{P}) \mathbf{W}^H + \mathbf{N}_t \quad (16)$$

where $\mathbf{H}_2 \in \mathbb{C}^{N \times N_{IRS}}$, $\mathbf{P} \in \mathbb{C}^{T \times N_{IRS}}$, $\mathbf{W} \in \mathbb{C}^{Q \times N_{IRS}}$, $\mathbf{Y}_t$ is the t -th matrix slice on the third-order tensor $\mathcal{Y} \in \mathbb{C}^{N \times Q \times T}$, $\mathbf{N}_t$ is the t -th matrix slice on the third-order tensor $\mathcal{N} \in \mathbb{C}^{N \times Q \times T}$. As Eq. (16) satisfies the requirements of the CP decomposition model, by the CP decomposition expression [15]-[17], Eq. (16) can be written as

$$\mathcal{Y}_{n,q,t} = \sum_{n_{IRS}=1}^{N_{IRS}} h_{2,n,n_{IRS}} w_{q,n_{IRS}} p_{t,n_{IRS}} + \mathcal{N}_{n,q,t} \quad (17)$$

where $\mathcal{Y}_{n,q,t} = [\mathcal{Y}]_{n,q,t}$, $\mathcal{N}_{n,q,t} = [\mathcal{N}]_{n,q,t}$, $h_{2,n,n_{IRS}} = [\mathbf{H}_2]_{n,n_{IRS}}$, $w_{q,n_{IRS}} = [\mathbf{W}]_{q,n_{IRS}}$, $p_{t,n_{IRS}} = [\mathbf{P}]_{t,n_{IRS}}$.

According to Eq. (17), we know that the received signal $\mathcal{Y}$ is a third-order tensor. Thus, it is difficult to estimate the channel $\mathbf{H}_1$ and the channel $\mathbf{H}_2$ directly. In such case, the third-order tensor needs to be rearranged to become a matrix, which is called matricization of tensor. In this paper, we use the model- n matrix unfolding method for the matricization of tensor.

Then the two types of matrices unfolding can be written as

$$\mathbf{Y}_{m1} = \mathbf{H}_2(\mathbf{P} \odot \mathbf{W})^H + \mathbf{N}_{m1} \quad (18)$$

$$\mathbf{Y}_{m2} = \mathbf{W}(\mathbf{H}_2 \odot \mathbf{P})^H + \mathbf{N}_{m2} \quad (19)$$

where $\mathbf{Y}_{m1} = [\mathbf{Y}_1, \dots, \mathbf{Y}_T] \in \mathbb{C}^{N \times QT}$ and $\mathbf{Y}_{m2} = [\mathbf{Y}_1^T, \dots, \mathbf{Y}_T^T] \in \mathbb{C}^{Q \times NT}$ are the mode-1 and model-2 matrices unfolding of the $\mathcal{Y}$, respectively. $\mathbf{N}_{m1} = [\mathbf{N}_1, \dots, \mathbf{N}_T] \in \mathbb{C}^{N \times QT}$ and $\mathbf{N}_{m2} = [\mathbf{N}_1^T, \dots, \mathbf{N}_T^T] \in \mathbb{C}^{Q \times NT}$ are the mode-1 and model-2 matrices unfolding of the $\mathcal{N}$, respectively. $\odot$ denotes the Khatri-Rao product.

Through Eq. (18) and Eq. (19), the third-order tensor problem is transformed to solve a problem with a second-order matrix, which reduces the complexity of analysis and calculation. At this time, the alternative least squares (ALS) method [18] can be used to estimate the channel $\mathbf{H}_1$ and the channel $\mathbf{H}_2$. But it is known that the operation of pseudo-inverse is required when the ALS method is used to solve the problem. Particularly, the divergence or slow convergence will occur when we solve the pseudo-inverse of the singular matrix. In order to avoid the above situation, this paper proposes a regular alternating least squares (RALS) scheme. The main concept of the RALS scheme is to regularize the cost function of alternating least squares. In doing so, the cost functions of regular alternating least squares can be reformulated as

$$J(\mathbf{H}_2) = \|\mathbf{Y}_{m1} - \mathbf{H}_2(\mathbf{P} \odot \mathbf{W})^H\|_F^2 + \sigma \|\mathbf{H}_2\|_F^2 \quad (20)$$

$$J(\mathbf{H}_1) = \|\mathbf{Y}_{m2} - (\mathbf{H}_1 \mathbf{X})^T (\mathbf{H}_2 \odot \mathbf{P})^H\|_F^2 + \delta \|\mathbf{H}_1\|_F^2 \quad (21)$$

where $\sigma \in \mathbb{R}$ and $\delta \in \mathbb{R}$ are the regularization parameters.

The estimation of the channel $\mathbf{H}_1$ and the channel $\mathbf{H}_2$ is to find the minimum value of the aforementioned cost functions.

Thus, the estimation of $\mathbf{H}_1$ and $\mathbf{H}_2$ can be respectively written as

$$\hat{\mathbf{H}}_2 = \arg \min_{\mathbf{H}_2} J(\mathbf{H}_2) \quad (22)$$

$$\hat{\mathbf{H}}_1 = \arg \min_{\mathbf{H}_1} J(\mathbf{H}_1) \quad (23)$$

The conjugate gradient of the cost function in Eq. (22) and Eq. (23) with respect to the variables of $\mathbf{H}_1$ and $\mathbf{H}_2$ will be solved

respectively. By making the conjugate gradient matrix equal to the zero matrix, the results can be written as

$$\begin{aligned} \frac{\partial J(\mathbf{H}_2)}{\partial \mathbf{H}_2^H} &= 0 \\ \mathbf{H}_2 \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I} \right] &= \mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) \end{aligned} \quad (24)$$

$$\Rightarrow \hat{\mathbf{H}}_2 = \mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I} \right]^{-1} \quad (25)$$

$$\begin{aligned} \frac{\partial J(\mathbf{H}_1)}{\partial \mathbf{H}_1^H} &= 0 \\ (\mathbf{X} \mathbf{H}_1)^H \left[(\mathbf{H}_2 \odot \mathbf{P})^H (\mathbf{H}_2 \odot \mathbf{P}) + \delta \mathbf{I} \right] &= \mathbf{Y}_{m2} (\mathbf{H}_2 \odot \mathbf{P}) \end{aligned} \quad (26)$$

$$\Rightarrow \hat{\mathbf{H}}_1 = \left\{ (\mathbf{X}^H)^{\dagger} \mathbf{Y}_{m2} (\mathbf{H}_2 \odot \mathbf{P}) \left[(\mathbf{H}_2 \odot \mathbf{P})^H (\mathbf{H}_2 \odot \mathbf{P}) + \delta \mathbf{I} \right]^{-1} \right\}^H \quad (27)$$

For the following comparative analysis of our proposed RALS scheme and the ALS scheme in [18], we focus on analyzing the estimation performance of the channel $\mathbf{H}_2$. The analysis of the channel $\mathbf{H}_1$ can be obtained similarly. The expression for estimating the channel $\mathbf{H}_2$ by the ALS scheme in [13] can be written as the following

$$\hat{\mathbf{H}}_{2-ALS} = \mathbf{Y}_{m1} \left[(\mathbf{P} \odot \mathbf{W})^H \right]^{\dagger} \quad (28)$$

In order to make the form of Eq. (28) and that of Eq. (25) be similar, Eq. (28) needs to be rewritten. Use the Moore-Penrose inverse matrix property of Khatri-Rao product and the relationship between Khatri-Rao product and Hadamard product, which can be written as [19]

$$\left[(\mathbf{P} \odot \mathbf{W})^H \right]^{\dagger} = (\mathbf{P} \odot \mathbf{W}) (\mathbf{P}^H \mathbf{P} * \mathbf{W}^H \mathbf{W})^{\dagger} \quad (29)$$

and

$$(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) = (\mathbf{P}^H \mathbf{P}) * (\mathbf{W}^H \mathbf{W}) \quad (30)$$

Then, Eq. (28) can be rewritten as

$$\begin{aligned}
\hat{\mathbf{H}}_{2-ALS} &= \mathbf{Y}_{m1} \left[(\mathbf{P} \odot \mathbf{W})^H \right]^\dagger \\
&= \mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) (\mathbf{P}^H \mathbf{P} * \mathbf{W}^H \mathbf{W})^\dagger \\
&= \mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \right]^\dagger
\end{aligned} \tag{31}$$

From Eq. (25) and Eq. (31), we can see that the pseudo-inverse matrix can be converted into the inverse matrix by adding the regularization parameter σ . The proposed RALS scheme is to add σ to each diagonal element of the singular matrix, which turns solving the pseudo-inverse of the singular matrix into solving the inverse of the nonsingular matrix.

3.1 Performance analysis of schemes

To compare the performance of the RALS scheme and the ALS scheme, we consider the two aspects, i.e., the singularity and the numerical stability. Then the mean square error (MSE) is derived to evaluate the channel estimation accuracy.

First, let us analyze the singular property, which comes from the pseudo-inverse of the singular matrix. The MSE of the RALS scheme can be written as

$$\begin{aligned}
MSE(\hat{\mathbf{H}}_2) &= \|\mathbf{H}_2 - \hat{\mathbf{H}}_2\|_F^2 \\
&\leq \|\mathbf{H}_2\|_F^2 + \|\hat{\mathbf{H}}_2\|_F^2 \\
&= \|\mathbf{H}_2\|_F^2 + \left\| \mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I} \right]^{-1} \right\|_F^2
\end{aligned} \tag{32}$$

If the matrix $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W})$ is a singular matrix, we need to ensure the matrix $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}$ needs to be a nonsingular matrix. We know that the determinant of a singular matrix is equal to zero. If $[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}] = 0$, the number of solutions of σ is limited. So there exists $r \in \mathbb{R}$, such that $[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}] \neq 0$ when $\sigma \geq r$. That is, by carefully designing σ , the matrix $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}$ is a nonsingular matrix. The design procedure of σ will be given in the next subsection. Let $\mathbf{P} \odot \mathbf{W} = \mathbf{U} \Sigma \mathbf{V}^H$ be the singular value decomposition of matrix $\mathbf{P} \odot \mathbf{W}$, where $\mathbf{U} \in \mathbb{C}^{TQ \times TQ}$ and $\mathbf{V} \in \mathbb{C}^{N_{IRS} \times N_{IRS}}$ are unitary matrices respectively, $\Sigma \in \mathbb{C}^{N_{IRS} \times N_{IRS}}$ is a diagonal matrix. Therefore, $MSE(\hat{\mathbf{H}}_2)$ can be rewritten as

$$\begin{aligned}
MSE(\hat{\mathbf{H}}_2) &\leq \|\mathbf{H}_2\|_F^2 + \left\| \mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I} \right]^{-1} \right\|_F^2 \\
&= \|\mathbf{H}_2\|_F^2 + \left\| \mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) \mathbf{V} (\Sigma^2 + \sigma \mathbf{I})^{-1} \mathbf{V}^H \right\|_F^2
\end{aligned} \tag{33}$$

where the range of each diagonal element of $(\Sigma^2 + \sigma \mathbf{I})^{-1}$ is $\left[(\sigma + \lambda_{\min})^{-1}, (\sigma + \lambda_{\max})^{-1} \right]$, $\lambda_{\max}$ is the maximum eigenvalue of $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W})$, $\lambda_{\min}$ is the minimum eigenvalue of $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W})$. It can be seen from Eq. (33) that the divergence

of the result of $\left\| \mathbf{Y}_{m1}(\mathbf{P} \odot \mathbf{W}) \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I} \right]^{-1} \right\|_F^2$ is avoided.

If the ALS scheme is used, there may exist one or more zero eigenvalues, since the rank of the matrix $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W})$ is deficient. Then the infinity appears in the diagonal elements of Σ^{-1} . Therefore,

$MSE(\hat{\mathbf{H}}_{2-ALS})$ can be written as

$$\begin{aligned} MSE(\hat{\mathbf{H}}_{2-ALS}) &= \left\| \mathbf{H}_2 - \hat{\mathbf{H}}_{2-ALS} \right\|_F^2 \\ &\leq \left\| \mathbf{H}_2 \right\|_F^2 + \left\| \mathbf{Y}_{m1}(\mathbf{P} \odot \mathbf{W}) \mathbf{V} \Sigma^{-2} \mathbf{V}^H \right\|_F^2 = \infty \end{aligned} \quad (34)$$

Therefore, by comparing $MSE(\hat{\mathbf{H}}_2)$ and $MSE(\hat{\mathbf{H}}_{2-ALS})$, we can see that adding regularization parameter can turn the singular matrix into the nonsingular matrix and avoid the divergence of the result when solving the inverse of the matrix. Thus, it can be concluded the proposed RALS scheme improves the performance of MSE.

Moreover, let's analyze the numerical stability of the proposed method. It is observed from Eq. (25) that whether the numerical calculation is stable lies on the inverse solution of the matrix $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}$. Therefore, the property of the matrix $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}$ needs to be analyzed.

Since the condition number is an important indicator for judging the numerical stability, we can determine whether a matrix is well-conditioned by calculating its condition number. The condition number of the matrix $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}$ and that of $(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W})$ without regularization parameter can be respectively expressed as

$$\begin{aligned} \text{cond} \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I} \right] &\leq \frac{\sigma + \left\| \mathbf{P} \odot \mathbf{W} \right\|_2^2}{\sigma} \\ &= 1 + \frac{\left\| \mathbf{P} \odot \mathbf{W} \right\|_2^2}{\sigma} \end{aligned} \quad (35)$$

$$\begin{aligned} \text{cond} \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \right] &= \left[\text{cond}(\mathbf{P} \odot \mathbf{W}) \right]^2 \\ &= \left(\frac{\left\| \mathbf{P} \odot \mathbf{W} \right\|_2}{\sqrt{\lambda_{\min}}} \right)^2 \leq \infty \end{aligned} \quad (36)$$

where $\sqrt{\lambda_{\min}}$ is the minimum singular value of $\mathbf{P} \odot \mathbf{W}$. By comparing the condition number

$\text{cond} \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I} \right]$ and $\text{cond} \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \right]$. We can see that adding regularization parameter can significantly improve matrix performance and turn the ill-conditioned matrix into a well-conditioned one. Therefore, the regularization parameter can be added to improve numerical stability.

In summary, adding regularization parameter can solve the problem of the singular matrix and the numerical stability of ill-conditioned matrix. Therefore, the proposed RALS scheme is better than the ALS scheme in terms of the MSE

performance and numerical stability.

3.2 Selection of the regularization parameter

By comparing the MSE performance of the RALS scheme and the ALS scheme, the superiority of the RALS scheme is demonstrated in the last subsection. In this subsection, we will analyze the selection of the regularization parameter. Here the analysis of the channel $\mathbf{H}_2$ is presented firstly, and the analysis of the channel $\mathbf{H}_1$ can also gotten in a similar way.

Here, $MSE(\hat{\mathbf{H}}_2) = \|\hat{\mathbf{H}}_2 - \mathbf{H}_2\|_F^2$ can be used to optimize the regularization parameter σ . Our purpose is to find the relationship between the regularization parameter σ and $\|\hat{\mathbf{H}}_2 - \mathbf{H}_2\|_F^2$. For the convenience of analysis, we let

$$\|\hat{\mathbf{H}}_2 - \mathbf{H}_2\|_F^2 = \|\hat{\mathbf{H}}_2 - \Theta + \Theta - \mathbf{H}_2\|_F^2 \quad (37)$$

where $\Theta = \bar{\mathbf{Y}}_{m1}(\mathbf{P} \odot \mathbf{W})[(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}]^{-1}$, and $\bar{\mathbf{Y}}_{m1} = \mathbf{H}_2(\mathbf{P} \odot \mathbf{W})^H$. According to the triangle inequality and the estimation of the channel $\mathbf{H}_2$ in Eq. (25), we can get the inequation as shown below

$$\begin{aligned} \|\hat{\mathbf{H}}_2 - \mathbf{H}_2\|_F^2 &\leq \|\hat{\mathbf{H}}_2 - \bar{\mathbf{Y}}_{m1}(\mathbf{P} \odot \mathbf{W})[(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}]^{-1}\|_F^2 \\ &\quad + \|\bar{\mathbf{Y}}_{m1}(\mathbf{P} \odot \mathbf{W})[(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}]^{-1} - \mathbf{H}_2\|_F^2 \\ &\leq \|\mathbf{Y}_{m1} - \bar{\mathbf{Y}}_{m1}\|_F^2 \left\| (\mathbf{P} \odot \mathbf{W})[(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}]^{-1} \right\|_F^2 \\ &\quad + \|\mathbf{H}_2(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W})[(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}]^{-1} - \mathbf{H}_2\|_F^2 \end{aligned} \quad (38)$$

Obviously, $\|\mathbf{Y}_{m1} - \bar{\mathbf{Y}}_{m1}\|_F^2 = \|\mathbf{N}_{m1}\|_F^2$, which represents the received noise. Through analyzing the above inequality, we can see

that the $\|\hat{\mathbf{H}}_2 - \mathbf{H}_2\|_F^2$ is divided into two parts. In the first part, since the value of $\|\mathbf{N}_{m1}\|_F$ is a fixed value, the value of the first

part depends on the value of $\left\| (\mathbf{P} \odot \mathbf{W})[(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}]^{-1} \right\|_F^2$. The $\left\| (\mathbf{P} \odot \mathbf{W})[(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}]^{-1} \right\|_F^2$ tends to

infinity as σ tends to zero and the matrix $(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W})$ isn't full rank. Therefore, from the point of view of numerical stability, σ should be as large as possible. While in the second part,

$\left\| \mathbf{H}_2(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W})[(\mathbf{P} \odot \mathbf{W})^H(\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I}]^{-1} - \mathbf{H}_2 \right\|_F^2$ tends to zero as σ tends to zero. According to the above analysis,

these two parts of the $MSE(\hat{\mathbf{H}}_2)$ are contradictory about the value of σ . In order to balance the two parts, it is required to choose a suitable regularization parameter σ .

According to the above analysis, our purpose is to choose a suitable regularization parameter σ , so that the minimum

value of the $MSE(\hat{\mathbf{H}}_2)$ (i.e., $\min \|\hat{\mathbf{H}}_2 - \mathbf{H}_2\|_F^2$) is obtained. And then, we can construct a MSE function expression about σ , which can be written as

$$\begin{aligned} f(\sigma) &= \left\| (\hat{\mathbf{H}}_2 - \mathbf{H}_2)(\mathbf{P} \odot \mathbf{W})^H - \mathbf{N}_{ml} \right\|_F^2 - \|\mathbf{N}_{ml}\|_F^2 \\ &= \left\| \hat{\mathbf{H}}_2 (\mathbf{P} \odot \mathbf{W})^H - \mathbf{Y}_{ml} \right\|_F^2 - \|\mathbf{N}_{ml}\|_F^2 \end{aligned} \quad (39)$$

In this case, the problem of solving the minimum value has become the problem of solving the zero point of $f(\sigma)$. First, we need to obtain the first derivative of the function in Eq. (39) with respect to the σ , which can be written as (Refer to APPENDIX A for the detailed derivation process)

$$f'(\sigma) = 2\sigma \left\| (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 + 2\sigma^2 \left\| \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 \quad (40)$$

where

$$\frac{d\hat{\mathbf{H}}_2}{d\sigma} = -\mathbf{Y}_{ml}(\mathbf{P} \odot \mathbf{W}) \left\{ \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I} \right]^{-1} \right\}^2 \quad (41)$$

It is obvious that $f'(\sigma) > 0$ when $\sigma > 0$, and thus $f(\sigma)$ is increasing monotonically with σ . And then, we need to obtain the second derivative of the function in Eq. (39) with respect to the σ , which can be written as (Refer to APPENDIX B for the detailed derivation process)

$$f''(\sigma) = 2 \left\| (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 - 4\sigma \left\| \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 \quad (42)$$

Letting $f''(\sigma) = 0$, we have

$$\sigma = \sigma_0 = \frac{\left\| (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2}{2 \left\| \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2} \quad (43)$$

From Eq. (42) and Eq. (43), it is obtained that $f''(\sigma) > 0$ for $\sigma \in (0, \sigma_0)$. Thus, $f(\sigma)$ is an increasing monotonically concave function when $\sigma \in (0, \sigma_0)$. Next, we need to find out a $\sigma \in (0, \sigma_0)$ to satisfy $f(\sigma) = 0$.

Since it satisfies the convergence condition of Newton iterative method, the Newton iterative method can be used to solve the above problem. The iterative expression can be written as

$$\sigma^{(n+1)} = \sigma^{(n)} - \frac{f(\sigma^{(n)})}{f'(\sigma^{(n)})} \quad (44)$$

where $\sigma^{(n)}$ is the n -th ($n \in \mathbb{N}$) iterative value of σ .

According to the above analysis, the method for determining σ is presented in Algorithm 1.

Algorithm1:Determine the regularization parameter σ

1. Set $n=0$, initialize $\sigma^{(n=0)} \in (0, \sigma_0)$
 2. Calculate $f(\sigma^{(n)}), f'(\sigma^{(n)})$
 3. $\sigma^{(n+1)} = \sigma^{(n)} - \frac{f(\sigma^{(n)})}{f'(\sigma^{(n)})}$
 4. Judge whether it is satisfied $|\sigma^{(n+1)} - \sigma^{(n)}| \leq 10^{-4}$. If it is satisfied, exit the iteration. Otherwise, go to the next step.
 5. Let $n=n+1$, repeat steps 2-4.
-

3.3 RALS algorithm

Since Eq. (27) involves the pseudo-inverse of $\mathbf{X}^H$, the pilot symbol matrix $\mathbf{X}$ is designed as the discrete Fourier transform (DFT) matrix to avoid the pseudo-inverse problem. The pilot symbol matrix $\mathbf{X}$ can be expressed as

$$\mathbf{X}_{q,m} = e^{-j \frac{2\pi(Q-1)(M-1)}{M}} \quad (45)$$

According to the properties of the DFT matrix [20], we can get $(\mathbf{X}^H)^\dagger = \frac{1}{Q} \mathbf{X}$. So Eq. (27) can be written as

$$\hat{\mathbf{H}}_1 = \left\{ \frac{1}{Q} \mathbf{X} \mathbf{Y}_{m2} (\mathbf{H}_2 \odot \mathbf{P}) [(\mathbf{H}_2 \odot \mathbf{P})^H (\mathbf{H}_2 \odot \mathbf{P}) + \delta \mathbf{I}]^{-1} \right\}^H \quad (46)$$

The selection of δ is based on Eq. (46). Due to the form of Eq. (46) is similar to Eq. (25), we also can construct a function expression $g(\delta)$ about δ , then δ can be obtained by the similar procedures in Algorithm 1. The summary of the proposed RALS algorithm is given in Algorithm 2.

Algorithm2:Regular alternating least squares_(RALS)

1. Set $i=0$, initialize $\mathbf{H}_1^{(i=0)}, \sigma^{(i=0)} \in (0, \sigma_0)$,
 $\delta^{(i=0)} \in (0, \sigma_0), e^{(i=0)}$
2. $i=i+1$
3. $\hat{\mathbf{H}}_2^{(i)} = \mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) [(\mathbf{P} \odot \mathbf{W}^{(i-1)})^H (\mathbf{P} \odot \mathbf{W}^{(i-1)}) + \sigma^{(i-1)} \mathbf{I}]^{-1}$
 $\hat{\mathbf{H}}_1^{(i)} = \left\{ \frac{1}{Q} \mathbf{X} \mathbf{Y}_{m2} (\mathbf{H}_2 \odot \mathbf{P}) [(\mathbf{H}_2^{(i)} \odot \mathbf{P})^H (\mathbf{H}_2^{(i)} \odot \mathbf{P}) + \delta^{(i-1)} \mathbf{I}]^{-1} \right\}^H$
4. Judge whether it is satisfied $\|e^{(i)} - e^{(i-1)}\| \leq 10^{-6}$, where $e^{(i)} = \|\mathcal{Y} - \mathcal{Y}^{(i)}\|_F^2$.

- If it is satisfied, exit the iteration. Otherwise go to the next step.
5. Solve $\sigma^{(i)}$ and $\delta^{(i)}$ by invoking Algorithm 1, repeat steps 2-4.
-

4 Method, results and discussions

In order to test the performance of the proposed scheme, the normalized mean square error (NMSE) is used to evaluate the channel estimation accuracy, and the expression is as follows:

$$NMSE(\hat{\mathbf{H}}) = \frac{1}{R} \sum_{r=1}^R \frac{\|\mathbf{H}_r - \hat{\mathbf{H}}_r\|_F^2}{\|\mathbf{H}_r\|_F^2} \quad (47)$$

where R represents the number of Monte Carlo rounds, and $\hat{\mathbf{H}}_r$ represents the result of the r -th complete channel estimation.

Assume $M=16$, $N=4$, $L_1=L_2=3$, $Q=20$, $T=50$, $R=5000$ in the simulations. To analyze the impact the number of IRS elements, it is assumed $N_{IRS} \in \{16, 25, 36, 49\}$.

Fig. 2 is a performance simulation diagram of the channel $\mathbf{H}_1$ and the channel $\mathbf{H}_2$ when the number of elements on the IRS is 16 and 49, respectively. It can be seen from Fig. 2 that the performance of the channel $\mathbf{H}_1$ and the channel $\mathbf{H}_2$ are both satisfactory. This is because the optimal regularization parameters are devised by using convex optimization theory, which is shown in Algorithm 1. Specifically, the cost functions in Eq. (20) and Eq. (21) are convex, then the designed method can converge to the global optimal solution.

Fig. 3 is a performance simulation diagram of the complete channel $\mathbf{H}$ when the number of elements on the IRS is 16, 25, 36, and 49 respectively. It can be seen from Fig. 2 and Fig. 3 that as the number of elements on the IRS increases, the NMSE performance of the proposed scheme is decreasing. This is because when the number of elements on the IRS increases, the channel coefficients that need to be estimated also increase, that is, the number of unknown items increases. To deal with this problem, the length of the pilot symbol sequence should be increased or the number of time frames should be increased.

Fig. 4 is a performance comparison diagram between the RALS algorithm proposed in this paper, the BALS (bilinear alternating least squares) algorithm in [12], and the conventional LS (least squares) algorithm. It can be seen from the simulations that the performance of the RALS algorithm is better than other comparison algorithms. On the one hand, the reason is that the conventional LS algorithm is used here as a cascade estimation. Although the conventional LS algorithm is simple and efficient, it ignores the influence of noise, resulting in a large NMSE at low signal noise ratio (SNR). On the other hand, the reason is that although the ALS algorithm considers the influence of noise, the two cost functions obtained are not necessarily convex functions. The final result may not converge or the optimal solution may not be obtained. The RALS algorithm proposed in this paper not only considers the influence of noise, but also adds regularization parameters to ensure convergence, so that the optimal solution can be obtained.

Fig. 2 NMSE of channel $\mathbf{H}_1$ and channel $\mathbf{H}_2$ vs. number of elements on IRS

Fig. 3 NMSE of channel $\mathbf{H}$ vs. number of elements on IRS

Fig. 4 NMSE performance comparison diagram of RALS, BALS and LS

5 Conclusion

This paper proposes a channel estimation scheme using regular alternating least squares algorithm for IRS assisted millimeter wave communications. Due to the introduction of regularization, the proposed scheme can avoid divergence and slow convergence in the channel estimation process, and also avoid numerical instability caused by singular value decomposition in the operation of pseudo-inverse. In order to obtain the optimal solution, the regularization parameters are devised by invoking the convex optimization theory. In simulations, the performance of the proposed scheme is compared with those of the existing algorithms. It is shown that the proposed algorithm is indeed superior to other existing algorithms in terms of estimation error performance.

APPENDIX A

$$\begin{aligned}
f'(\sigma) &= \frac{d \operatorname{tr} \left\{ \left[\hat{\mathbf{H}}_2 (\mathbf{P} \odot \mathbf{W})^H - \mathbf{Y}_{ml} \right]^H \left[\hat{\mathbf{H}}_2 (\mathbf{P} \odot \mathbf{W})^H - \mathbf{Y}_{ml} \right] \right\}}{d\sigma} \\
&= \operatorname{tr} \left\{ \frac{d \left[\hat{\mathbf{H}}_2 (\mathbf{P} \odot \mathbf{W})^H - \mathbf{Y}_{ml} \right]^H \left[\hat{\mathbf{H}}_2 (\mathbf{P} \odot \mathbf{W})^H - \mathbf{Y}_{ml} \right]}{d\sigma} \right\} \\
&= \operatorname{tr} \left\{ \left[\frac{d\hat{\mathbf{H}}_2}{d\sigma} (\mathbf{P} \odot \mathbf{W})^H \right]^H \left[\hat{\mathbf{H}}_2 (\mathbf{P} \odot \mathbf{W})^H - \mathbf{Y}_{ml} \right] \right. \\
&\quad \left. + \left[\hat{\mathbf{H}}_2 (\mathbf{P} \odot \mathbf{W})^H - \mathbf{Y}_{ml} \right]^H \left[\frac{d\hat{\mathbf{H}}_2}{d\sigma} (\mathbf{P} \odot \mathbf{W})^H \right] \right\} \\
&= \operatorname{tr} \left\{ \left[\frac{d\hat{\mathbf{H}}_2}{d\sigma} (\mathbf{P} \odot \mathbf{W})^H \right]^H \left[\hat{\mathbf{H}}_2 (\mathbf{P} \odot \mathbf{W})^H - \mathbf{Y}_{ml} \right] \right\} \\
&\quad + \operatorname{tr} \left\{ \left[\hat{\mathbf{H}}_2 (\mathbf{P} \odot \mathbf{W})^H - \mathbf{Y}_{ml} \right]^H \left[\frac{d\hat{\mathbf{H}}_2}{d\sigma} (\mathbf{P} \odot \mathbf{W})^H \right] \right\} \\
&= \operatorname{tr} \left\{ \left(\frac{d\hat{\mathbf{H}}_2}{d\sigma} \right)^H \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \hat{\mathbf{H}}_2 - \mathbf{Y}_{ml} (\mathbf{P} \odot \mathbf{W}) \right] \right\} \\
&\quad + \operatorname{tr} \left\{ \frac{d\hat{\mathbf{H}}_2}{d\sigma} \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \hat{\mathbf{H}}_2 - \mathbf{Y}_{ml} (\mathbf{P} \odot \mathbf{W}) \right]^H \right\} \\
&= \operatorname{tr} \left[\left(\frac{d\hat{\mathbf{H}}_2}{d\sigma} \right)^H (-\sigma \hat{\mathbf{H}}_2) \right] + \operatorname{tr} \left[(-\sigma \hat{\mathbf{H}}_2)^H \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right] \\
&= 2 \operatorname{tr} \left[\sigma \left(\frac{d\hat{\mathbf{H}}_2}{d\sigma} \right)^H (\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} + \sigma^2 \left(\frac{d\hat{\mathbf{H}}_2}{d\sigma} \right)^H \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right] \\
&= 2\sigma \operatorname{tr} \left[\left(\frac{d\hat{\mathbf{H}}_2}{d\sigma} \right)^H (\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right] + 2\sigma^2 \operatorname{tr} \left[\left(\frac{d\hat{\mathbf{H}}_2}{d\sigma} \right)^H \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right] \\
&= 2\sigma \left\| (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 + 2\sigma^2 \left\| \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2
\end{aligned} \tag{48}$$

where

$$\begin{aligned}
& (\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \hat{\mathbf{H}}_2 + \sigma \hat{\mathbf{H}}_2 = \mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) \\
& \Rightarrow (\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} + \sigma \frac{d\hat{\mathbf{H}}_2}{d\sigma} = -\hat{\mathbf{H}}_2 \\
& \Rightarrow \frac{d\hat{\mathbf{H}}_2}{d\sigma} = -\mathbf{Y}_{m1} (\mathbf{P} \odot \mathbf{W}) \left\{ \left[(\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \mathbf{I} \right]^{-1} \right\}^2
\end{aligned} \tag{49}$$

APPENDIX B

$$\begin{aligned}
f''(\sigma) &= \frac{d \left[2\sigma \left\| (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 + 2\sigma^2 \left\| \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 \right]}{d\sigma} \\
&= 2 \left\| (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 + 4\sigma \text{tr} \left[\left(\frac{d^2 \hat{\mathbf{H}}_2}{d\sigma^2} \right)^H (\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right] \\
&\quad + 4\sigma \left\| \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 + 4\sigma^2 \text{tr} \left[\left(\frac{d^2 \hat{\mathbf{H}}_2}{d\sigma^2} \right)^H \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right] \\
&= 2 \left\| (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 + 4\sigma \left\| \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 \\
&\quad + 4\sigma \text{tr} \left\{ \left[\left(\frac{d^2 \hat{\mathbf{H}}_2}{d\sigma^2} \right)^H (\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) + \sigma \left(\frac{d^2 \hat{\mathbf{H}}_2}{d\sigma^2} \right)^H \right] \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\} \\
&= 2 \left\| (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 + 4\sigma \left\| \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 + 4\sigma \text{tr} \left[-2 \left(\frac{d\hat{\mathbf{H}}_2}{d\sigma} \right)^H \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right] \\
&= 2 \left\| (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2 - 4\sigma \left\| \frac{d\hat{\mathbf{H}}_2}{d\sigma} \right\|_F^2
\end{aligned} \tag{50}$$

where

$$\begin{aligned}
& (\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \frac{d\hat{\mathbf{H}}_2}{d\sigma} + \sigma \frac{d\hat{\mathbf{H}}_2}{d\sigma} = -\hat{\mathbf{H}}_2 \\
& \Rightarrow (\mathbf{P} \odot \mathbf{W})^H (\mathbf{P} \odot \mathbf{W}) \frac{d^2 \hat{\mathbf{H}}_2}{d\sigma^2} + \sigma \frac{d^2 \hat{\mathbf{H}}_2}{d\sigma^2} = -2 \frac{d\hat{\mathbf{H}}_2}{d\sigma}
\end{aligned} \tag{51}$$

Abbreviations

IRS: Intelligent reflecting surface; CSI: channel state information; RALS: Regular alternating least squares; ALS: Alternate least squares; MIMO: Multiple input multiple output; OFDM: Orthogonal frequency division multiplexing; NOMA: Non-orthogonal multiple access; BS: Base station; MS: Mobile station; CS: Compressed sensing; DL: Deep learning; AWGN: Additive white Gaussian noise; MSE: mean square error; NMSE: Normalized mean square error; BALS: Bilinear alternating least squares; LS: Least squares; SNR: signal noise ratio

Acknowledgements

This work was supported by the National and Local Joint Engineering Laboratory of RF Integration and Micro-Assembly Technology

Authors' contributions

CD proposes the innovation ideas and carries out experiments and data analysis. FYH, LC and WH reviewed the manuscript and fixed some issues. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grants 61771257 and 62171237.

Availability of data and materials

Not applicable

Declarations

Competing interests

The authors declare that they have no competing interests.

REFERENCES

1. K. Hassan, M. Masarra, M. Zwingelstein and I. Dayoub, "Channel Estimation Techniques for Millimeter-Wave Communication Systems: Achievements and Challenges," in *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1336-1363, 2020.
2. Q. Wu and R. Zhang, "Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming," in *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394-5409, Nov. 2019.
3. Q. Wu and R. Zhang, "Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network," in *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106-112, January 2020.
4. E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. -S. Alouini and R. Zhang, "Wireless Communications Through Reconfigurable Intelligent Surfaces," in *IEEE Access*, vol. 7, pp. 116753-116773, 2019.
5. C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A New Wireless Communication Paradigm through Software-Controlled Metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162-169, Sep 2018.
6. X. Yuan, Y.-J. Zhang, Y. Shi, W. Yan, and H. Liu, "Reconfigurable- Intelligent-Surface Empowered 6G Wireless Communications: Challenges and Opportunities," 2020,[Online]. Available: <https://arxiv.org/abs/2001.00364>.
7. B. Zheng and R. Zhang, "Intelligent Reflecting Surface-Enhanced OFDM: Channel Estimation and Reflection Optimization," in *IEEE Wireless Communications Letters*, vol. 9, no. 4, pp. 518-522, April 2020.
8. Y. Yang, B. Zheng, S. Zhang and R. Zhang, "Intelligent Reflecting Surface Meets OFDM: Protocol Design and Rate Maximization," in *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4522-4535, July 2020.
9. B. Zheng, Q. Wu and R. Zhang, "Intelligent Reflecting Surface-Assisted Multiple Access With User Pairing: NOMA or OMA?," in *IEEE Communications Letters*, vol. 24, no. 4, pp. 753-757, April 2020.
10. G. Yang, X. Xu and Y. Liang, "Intelligent Reflecting Surface Assisted Non-Orthogonal Multiple Access," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1-6, 2020.
11. A. Taha, M. Alrabeiah and A. Alkhateeb, "Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning," in *IEEE Access*, vol. 9, pp. 44304-44321, 2021.
12. G. T. de Araújo and A. L. F. de Almeida, "PARAFAC-Based Channel Estimation for Intelligent Reflective Surface Assisted MIMO System," *2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 1-5, 2020.
13. X. Ma, Z. Chen, Y. Chi, W. Chen, L. Du and Z. Li, "Channel Estimation for Intelligent Reflecting Surface Enabled Terahertz MIMO Systems," *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1-6, 2020.
14. Z. He and X. Yuan, "Cascaded Channel Estimation for Large Intelligent Metasurface Assisted Massive MIMO," in *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 210-214, Feb. 2020.
15. R. B. Cattell, "Parallel proportional profiles and other principles for determining the choice of factors by rotation. Psychometrika," 1944, 9: 267-283.

16. R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.
17. J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, 35 (1970), pp. 283–319.
18. E. Acar and B. Yener, "Unsupervised multiway data analysis: A literature survey," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 1, pp. 6-20, 2009.
19. X. D. Zhang. *Matrix Analysis and Application*[M]. Beijing: Tsinghua University Press, 2011, pp. 68-74.
20. Q. -U. -A. Nadeem, H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah and M. -S. Alouini, "Intelligent Reflecting Surface-Assisted Multi-User MISO Communication: Channel Estimation and Beamforming Design," in *IEEE Open Journal of the Communications Society*, vol. 1, pp. 661-680, 2020.

Figure Title and Legend

Figure 1 IRS assisted millimeter wave communication system

Legend: Figure 1 is an IRS assisted millimeter wave communication system, the communication system has two propagation links, which are the BS directly to the MS and the BS to the MS through the IRS. There are three kinds of channels on these two links: the channel from the BS directly to the MS $\mathbf{L}$, the channel from the BS to the IRS $\mathbf{H}_1$, and the channel from the IRS to the MS $\mathbf{H}_2$. This article assumes that the direct path is blocked by the obstacles.

Figure 2 NMSE of channel $\mathbf{H}_1$ and channel $\mathbf{H}_2$ vs. number of elements on IRS

Legend: Figure 2 compares the performance of the normalized mean square error for the channel $\mathbf{H}_1$ and the channel $\mathbf{H}_2$ when the number of IRS elements are 16 and 49 respectively.

Figure 3 NMSE of channel $\mathbf{H}$ vs. number of elements on IRS

Legend: In order to analyze the influence of the number of IRS elements on the channel estimation, compare the normalized mean square error performance of the estimated channel $\mathbf{H}$ when the number of elements on the IRS are 16, 25, 36 and 49 respectively.

Figure 4 NMSE performance comparison diagram of RALS, BALS and LS

Legend: In order to prove the performance superiority of the proposed RALS algorithm, it is compared with the existing BALS algorithm and LS algorithm.