Unsupervised Channel Estimation with Dual Path Knowledge-Aware Auto-Encoder

Zhiheng Guo*, Yuanzhang Xiao† and Xiang Chen*‡
*School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China
guozhh7@mail2.sysu.edu.cn; chenxiang@mail.sysu.edu.cn
†Hawaii Advanced Wireless Technologies Institute, University of Hawaii, Honolulu, HI
yxiao8@hawaii.edu

***Abstract*—This paper proposes an unsupervised channel estimation method for massive MIMO systems. Our method builds on a novel improvement of the traditional Auto-Encoder (AE) by incorporating the knowledge of signal propagation models into the decoder. In our proposed knowledge-aware AE, instead of having learnable parameters, the decoder has fixed weights that implement the signal propagation model. Such modification forces the encoder to output meaningful physical parameters of interests (i.e., angle-of-arrivals, path gains and path angles), which cannot be achieved by standard AE. This paper rigorously analyzes the multiplicity of global optima in unsupervised channel estimation problems. Our analysis informs the design of the encoder as a dual-path neural network, which uses the received signal and its correlation matrix to estimate the path gains and the angle-of-arrivals, respectively. In the training phase, we update the parameters in the dual paths alternatively to alleviate the issue of multiple global optima. To further improve the performance, we propose an efficient method to compute good initial points for training. Numerical simulation results corroborate the analysis, and demonstrate the performance improvements of the proposed method over traditional channel estimation methods.**

I. Introduction

With the popularity of fifth generation (5G) communication networks, massive MIMO becomes one of the vital techniques to meet wireless communication system requirements such as explosive data traffic growth, high quality of service, increased energy efficiency requirement, etc. Theoretically, massive MIMO can easily enhance the capacity of a communication system with additional antennas [1]. However, there are some difficulties in the application of massive MIMO, including the sophisticated channel modeling, [2] the high cost of channel state information (CSI) [3], the huge amount of calculation due to the super-large matrix generated by massive MIMO, etc. As a result some effective schemes are urgently necessary to overcome these difficulties.

Generally, there are three ways to achieve channel and angle of arrival(AoA) estimation in Massive MIMO: direct estimation from the received signal [4], estimation based on compressed sensing (CS) [5] and deep learning-based estimation [6].

The estimation from the received CSI matrix and subspace-based estimation are two meaningful direct estimation methods from the received signal in the field of Massive MIMO. However, with the increase of the number of the antenna, the cost of CSI matrix becomes unbearable and both subspace-based estimation become computationally expensive. To avoid huge amount of calculation and the payload of CSI matrix, CS-based estimation effectively compresses the CSI transmission matrix by exploiting the potential sparsity of the Massive MIMO CSI in the certain transform domain. However, most CS-based channel estimation methods require strong sparsity channel which make it hardly work well when the channel has correlated path likes Rayleigh channel. In recent years, more and more schemes combining deep learning (DL) and channel estimation are proposed [7]. Combining with DL, the channel estimation methods become model-free that they do not have a strong assumption on channel characteristic and have low computational complexity after the model is trained. In [8], by leveraging the spatial structure, they integrated DL technology into the massive MIMO system and first proposed the use of DL to achieve AoA estimation and channel estimation based on deep neural network (DNN). However, to the best of our knowledge, there is still no one combines unsupervised network with channel estimation and AoA estimation in Massive MIMO scenarios.

There are two seemingly unbreakable bottlenecks between unsupervised network and channel and AoA estimation. One is that without the constraints of the estimated features labels, the channel features extracted from the unsupervised network becomes physical meaningless. Another is that without the labels, the network can only take the difference between input, received signal, and output signals, recovered from the estimated feature, as its loss function. As a result, the features estimated from the network will only be the representative number extracted from the received signal but may not be trend to the actual channel features. Moreover, because of the sophisticated channel model, there may be a lot of local minimum in the difference between received and recovery signal which will bring local convergence points in the network, also called multiplicity of global optima in this paper. Thus, network can hardly make an accurate estimation.

To our best knowledge, this paper first combines the unsupervised model-driven network based on Auto-Encoder(AE) [9] with channel estimation and AoA estimation in the field of Massive MIMO. By replacing the decoder with known channel model, this paper builds a Knowledge-Aware Auto-

‡ Xiang Chen is the corresponding author.

Encoder (KA-AE) network, which makes the output of encoder becomes physical meaningful. This paper also derives the loss function of the KA-AE and finds the convergence point which guides the design of the network. According to the derivation of loss function, this paper imposes the two phase training process to KA-AE, initialization and normal training. They will ensure the convergence of KA-AE and enhance the estimation accuracy. Furthermore, this paper uses dual path framework to design the encoder of the network for AoA estimation and channel estimation, respectively. Thus, the network implements greedy algorithm and becomes interpretable; Moreover, Under the guidance of the derived mathematical formula of loss function, this paper introduces the correlation codebook to eliminate multi-users interference and imposes sectorization method to overcome mutiplicity global optima problem. At last, the numerical simulation shows that this paper realizes high accurate channel estimation and AoA estimation whose performance is superior to MUSIC and tend to the supervised network.

II. Channel Model and Problem Formulation

Consider a massive MIMO uplink system consisting of one base station with a uniform linear array of N_t antennas and K users. The received signals at the base station during M snapshots can be represented as follows:

$$\boldsymbol{Y} = \boldsymbol{A}\left(\boldsymbol{H} \odot \boldsymbol{S}\right) + \boldsymbol{N}, \tag{1}$$

where $\odot$ is the Hadamard product, $\boldsymbol{Y} \in \mathbb{C}^{N_t \times M}$ is the collection of received signals during M snapshots, $\boldsymbol{A} \in \mathbb{C}^{N_t \times K}$ is the array response matrix, $\boldsymbol{H} \in \mathbb{C}^{K \times M}$ is the channel fading matrix, $\boldsymbol{S} \in \mathbb{C}^{K \times M}$ is the collection of K transmit signals during M snapshots, and $\boldsymbol{N} \in \mathbb{C}^{N_t \times M}$ is the Gaussian noise.

Each user k's location is specified by the distance to the base station and the AoA, defined as the angle of user k's impinging signal relative to the broadside of the antenna array (i.e., the line perpendicular to the antenna array). We assume that the users are stationary during the M snapshots. Therefore, user k's AoA $\theta_k \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ stays the same during the M snapshots. Given the AoAs, the array response matrix can be written as $\boldsymbol{A} = \left[\boldsymbol{a}\left(\theta_1\right), \boldsymbol{a}\left(\theta_2\right), \ldots, \boldsymbol{a}\left(\theta_K\right)\right]$ with

$$\boldsymbol{a}(\theta_k) = \left[1, e^{-j\frac{2\pi d}{\lambda}\sin\theta_k}, \ldots, e^{-j\frac{2\pi d}{\lambda}(N_t-1)\sin\theta_k}\right]^T,$$

where $j = \sqrt{-1}$, λ is the wavelength and $d \geq \frac{\lambda}{2}$ is the distance between adjacent antenna elements.

The distances from users to the base station, along with shadowing and small-scale fading, determine the channel fading matrix $\boldsymbol{H}$. The channel gain from user k to the base station in snapshot m is $[\boldsymbol{H}]_{k,m} = \alpha_{k,m} e^{j\phi_{k,m}}$, where $\alpha_{k,m}$ and $\phi_{k,m}$ are the magnitude and the phase of the complex channel gain. Unlike the AoAs, the path gains $\alpha_{k,m}$ and path angles $\phi_{k,m}$ are time varying during the M snapshots due to random fading effects.

In this paper, we aim to estimate the AoAs $\theta_1, \ldots, \theta_K$ and the path gains $\alpha_{k,m}$ and path angles $\phi_{k,m}$ from the received signals $\boldsymbol{Y}$ during the M snapshots.

III. Proposed Solution

This paper proposes an unsupervised channel estimation method by modifying the AE (see Fig. 1). The encoder takes the received signals and the self-correlation matrix as input, and outputs the parameters to estimate. The decoder reconstructs the received signal based on the output of the encoder. The key difference from the traditional AE is that *the decoder of the proposed network is fixed*, as opposed to have learnable parameters. As the model-driven network [10], the signal model is known as in (1), and is utilized in our design of the decoder.

The encoder of the proposed KA-AE network are divided into two parts. The upper one is the channel estimation network which will give the estimation of path gains $|\widehat{\alpha}|$ and path angles $\widehat{\phi}$ while another is the AoA estimation network whose output is the estimation of AoAs $\widehat{\theta}$. The channel estimation network takes the real and imaginary parts of the received signal matrix, called the target signal hereafter, as the $2N_t \times M$-dimensional input. The real and imaginary parts of the self-correlation of the received signal matrix $\boldsymbol{R}_{yy}$ is the input of the AoA estimation network.

The divide of the encoder make the network mimic the iteration process which will implement the greedy algorithm. The Codebook Correlation model of the proposed KA-AE network will provide the pseudo-label $\widetilde{\theta}$ for AoA estimation which will be used in initialization process. The decoder of the KA-AE network is replaced with the known signal model. Thus the decoder can recover the received signal matrix from the estimation output of the encoder. The training process of the proposed KA-AE network is divided into two phase, the initialization and the normal training of the network. With the pseudo-label $\widetilde{\theta}$, the initialization process, dotted lines in the Fig. 1, has an inaccurate supervision with the loss function below:

$$loss^{(init)} = \mathbb{E}\left[\left(\boldsymbol{Y}_R - \widetilde{\boldsymbol{Y}}_{\boldsymbol{R}}\right)^2 + \left(\boldsymbol{Y}_I - \widetilde{\boldsymbol{Y}}_{\boldsymbol{I}}\right)^2 + \left(\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}\right)^2\right]. \tag{2}$$

where $\widetilde{\boldsymbol{Y}}_{\boldsymbol{R}}$ and $\widetilde{\boldsymbol{Y}}_{\boldsymbol{I}}$ are the real and the imaginary part of the recovery signal by pseudo-label. However, with the lack of any label, the normal training can only take the Mean Squared Error (MSE) between the recovery signal and the target signal as the loss function:

$$loss^{(train)} = \mathbb{E}\left[\left\|\boldsymbol{Y} - \widehat{\boldsymbol{Y}}\right\|_F^2\right], \tag{3}$$

where $\|\cdot\|_F$ is the Frobenius norm, and $\mathbb{E}(\cdot)$ is the expectation operator. Here the expectation is taken over the parameters to estimate and the noise.

IV. Performance Analysis

A. Convergence Analysis of Loss Function

For the reason that the learnable parameters of the proposed KA-AE network is upgraded by the back propagation with the derivative of the loss function, the proposed KA-AE network will converge into the direction which the loss function gets smaller. For the normal training process , it

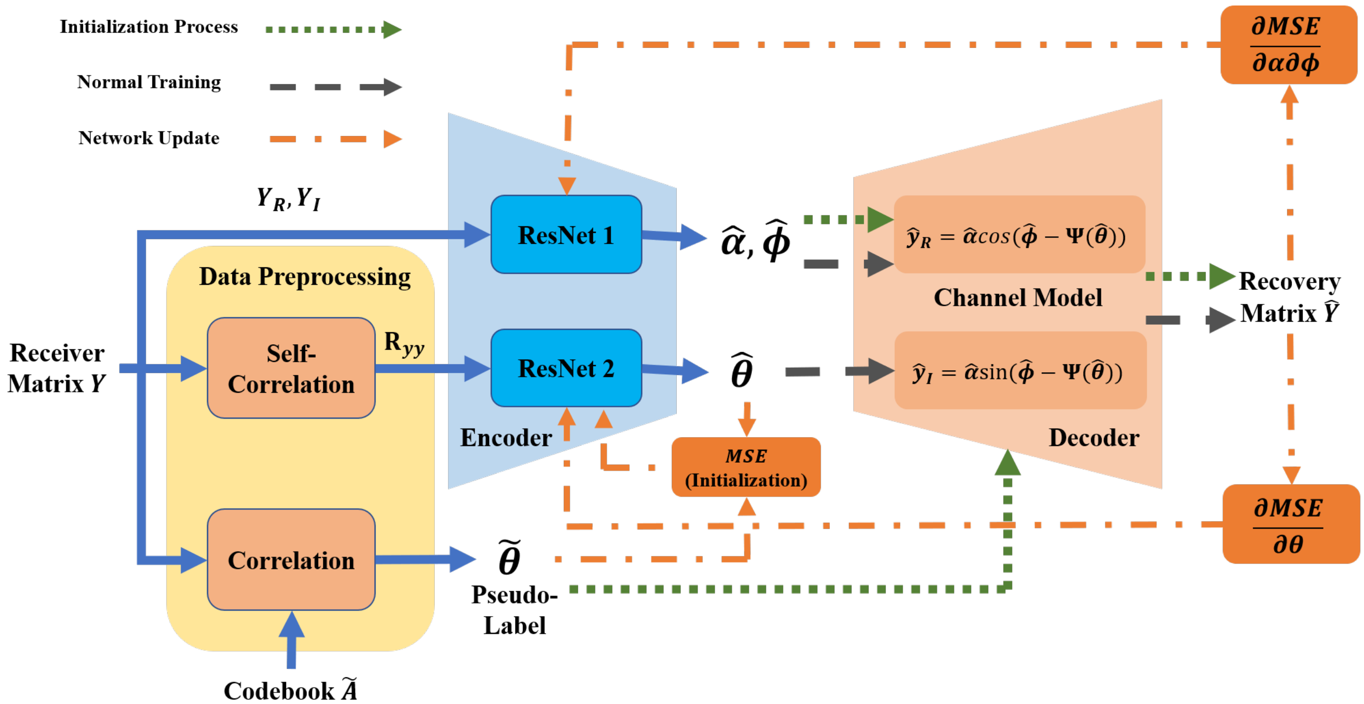


Fig. 1: The KA-AE network is composed of data preprocessing, encoder and decoder. The encoder consists of channel estimation ResNets and AoA estimation ResNet, while the decoder is the known channel model. The training process is divided into initialization process and normal training. The pseudo-label only takes part in the initialization process.

can be easily find that the loss function (3) is the MSE between the recovery signal matrix and the received signal matrix. When $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$, $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$ and $\widehat{\boldsymbol{\phi}} = \boldsymbol{\theta}$, the loss function will reach its minimum, i.e. $loss^{train} = 0$. To gain more insight, we derive a more explicit expression of the loss function in (4), where $\boldsymbol{y}_{m,k}^R = \alpha_{m,k} \cos\left(\phi_{m,k} - \boldsymbol{\Psi}\left(\theta_k\right)\right)$ and $\boldsymbol{y}_{m,k}^I = \alpha_{m,k} \sin\left(\phi_{m,k} - \boldsymbol{\Psi}\left(\theta_k\right)\right)$. $\widehat{\boldsymbol{y}}_{m,k}^R$ and $\widehat{\boldsymbol{y}}_{m,k}^I$ can be calculated by replacing α, ϕ, θ in $\boldsymbol{y}_{m,k}^R$ and $\boldsymbol{y}_{m,k}^I$ with $\widehat{\alpha}, \widehat{\phi}, \widehat{\theta}$.

From (4), we can see that the training loss (3) consists of two parts – one affected by the accuracy of the estimation of AoAs, path gains, and path angles, and the other one resulting from interference between different users.

$$\mathbb{E}\left[\boldsymbol{y}_{m,k}^R \boldsymbol{y}_{m,i}^R\right] = \alpha_{m,k}\alpha_{m,i}\mathbb{E}\left[\cos\boldsymbol{\beta}_{m,k}\cos\boldsymbol{\beta}_{m,i}\right]. \quad (5)$$

where $\boldsymbol{\beta}_{m,k} = \phi_{m,k} - \boldsymbol{\Psi}\left(\theta_k\right)$. Therefore, when $\boldsymbol{\beta}_{m,k} \neq \boldsymbol{\beta}_{m,i}$, there will be $\mathbb{E}\left[\boldsymbol{y}_{m,k}^R \boldsymbol{y}_{m,i}^R\right] = 0$. For the reason that in the fields of AoAs estimation, different users with the same AoA cannot be distinguished, this paper makes an assumption:

Assumption 4.1: There is no different users with the same AoA, i.e. $\theta_k \neq \theta_i, i \neq k$. What's more, for $\theta \in [-\pi/2, \pi/2]$, $\boldsymbol{\Psi}\left(\theta_k\right) \neq \boldsymbol{\Psi}\left(\theta_j\right), i \neq k$, which means:

$$\mathbb{E}\left[\cos\boldsymbol{\beta}_{m,k}\cos\boldsymbol{\beta}_{m,i}\right] = 0, i \neq k. \quad (6)$$

Thus, it can be easily find out that if $\widehat{\theta}_k \to \theta_k, \widehat{\theta}_j \to \theta_j$ and $\theta_k \neq \theta_j, i \neq k$, there are $\mathbb{E}\left[\widehat{\boldsymbol{y}}_{m,k}^R \boldsymbol{y}_{m,i}^R\right] \to 0, \mathbb{E}\left[\boldsymbol{y}_{m,k}^R \widehat{\boldsymbol{y}}_{m,i}^R\right] \to 0$ and $\mathbb{E}\left[\widehat{\boldsymbol{y}}_{m,k}^R \widehat{\boldsymbol{y}}_{m,i}^R\right] \to 0$. Thus, the later term of the training loss function will not affect the convergence of the proposed KA-AE network, which means there is no more interference between different users during convergence.

B. Initialization Based on Correlation Codebook

In order to avoid the randomness of the network at the beginning of the training which will affect the convergence of the loss function, this paper divides the network training

$$\begin{aligned} loss^{(train)} = \; & \mathbb{E}\left[\sum_{m=0}^{M-1}\sum_{k=0}^{K-1}\left(\alpha_{m,k}^2 + \widehat{\alpha}_{m,k}^2 - 2\alpha_{m,k}\widehat{\alpha}_{m,k}\cos\left(\left(\phi_{m,k} - \widehat{\phi}_{m,k}\right) - \left(\boldsymbol{\Psi}\left(\theta_k\right) - \boldsymbol{\Psi}\left(\widehat{\theta}_k\right)\right)\right)\right)\right], \\ & + \mathbb{E}\left[\sum_{m=0}^{M-1}\sum_{k=0}^{K-1}\sum_{\substack{i=0,\\ i\neq k}}^{K-1}\left(\left(\boldsymbol{y}_{m,k}^R - \widehat{\boldsymbol{y}}_{m,k}^R\right)\left(\boldsymbol{y}_{m,i}^R - \widehat{\boldsymbol{y}}_{m,i}^R\right) + \left(\boldsymbol{y}_{m,k}^I - \widehat{\boldsymbol{y}}_{m,k}^I\right)\left(\boldsymbol{y}_{m,i}^I - \widehat{\boldsymbol{y}}_{m,i}^I\right)\right)\right]. \end{aligned} \quad (4)$$

process into two phase, initialization and normal training. The initialization process is to make $\widehat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}$ before normal training.

The initialization process builds a codebook to provide the pseudo-label for the proposed KA-AE network. Specifically, we define the correlation codebook as $\overline{\boldsymbol{A}} = \left[e^{-j\boldsymbol{\Psi}(\theta_{min})}, e^{-j\boldsymbol{\Psi}(\theta_{min}+\Delta\theta)}, \ldots, e^{-j\boldsymbol{\Psi}(\theta_{max})}\right]$, where θ_{min} is the minimum AoA, θ_{max} is the maximum AoA, and $\Delta\theta$ is the precision of the initialization. Then we calculate the correlation between the received signal matrix and the codebook $\overline{\boldsymbol{c}}_{\boldsymbol{YA}} = \boldsymbol{1}^T\boldsymbol{Y}\overline{\boldsymbol{A}}^H$, where $\boldsymbol{1} \in \mathbb{R}^{N_t}$ is the vector of all ones. The initial $\widetilde{\boldsymbol{\theta}}$ are the K angles that result in the K largest values in $\overline{\boldsymbol{c}}_{\boldsymbol{YA}}$.

The proposed KA-AE network uses the initial $\widetilde{\boldsymbol{\theta}}$ in the loss function (2) during the initialization process. Thus, we have $\widehat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}$ before the normal training.

C. Multiplicity of Global Optima

According to *Assumption* 4.1, if $\widehat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}$ is hold, the last term of the training loss function (4) is tend to 0. Thus,

$$loss = \mathbb{E}\left[\sum_{m=0}^{M-1}\sum_{k=0}^{K-1}\left(\alpha_{m,k}^2 + \widehat{\alpha}_{m,k}^2 - 2\alpha_{m,k}\widehat{\alpha}_{m,k}\cos\boldsymbol{\eta}_{\phi,\theta}^{(m,k)}\right)\right], \quad (7)$$

where $\boldsymbol{\eta}_{\phi,\theta}^{(m,k)} = \left(\phi_{m,k} - \widehat{\phi}_{m,k}\right) - \left(\boldsymbol{\Psi}(\theta_k) - \boldsymbol{\Psi}\left(\widehat{\theta}_k\right)\right)$. As shown in (7), there is not inter-users interference in the KA-AE network. The muti-users estimation fallback to the stacking of multiple single-user estimation problems with initialization. However, as the (7) shown, there are three variable $\widehat{\alpha}, \widehat{\phi}, \widehat{\theta}$ in the loss function and $\widehat{\phi}$ and $\widehat{\theta}$ interfere with each other in the same cosine. The derivation of the global convergence of loss function (7) becomes intractable.

In order to analyze the global convergence of the KA-AE network, this paper divides the encoder into two parts, the channel estimation network for path gain $\widehat{\boldsymbol{\alpha}}$ and pain angle $\widehat{\boldsymbol{\phi}}$ and the AoAs estimation network for AoAs $\widehat{\boldsymbol{\theta}}$. Thus, when the parameters of the channel estimation network is updating, the estimation of AoAs $\widehat{\boldsymbol{\theta}}$ can be regarded as a constant which is similar to the AoAs estimation network. Then the global convergence of each network can be derived separately.

Proposition 1: Consider there is no interference between difference users. If $\overline{\theta} \rightarrow \theta$ is hold, the convergence point of the channel estimation network must hold the proposition that $\widehat{\phi} \rightarrow \phi + 2\pi l, (\phi, \widehat{\phi}) \in (-\pi, \pi] \quad l \in Z$ and $\widehat{\alpha} \rightarrow \alpha$. It means that the KA-AE has an accurate channel estimation.

Proof: Please see our online appendix [11] for the complete proof. ■

Thus, with initialization, $\overline{\theta} \rightarrow \theta$ is hold and the channel estimation network will converge into the point where the estimated path gain $\widehat{\alpha}$ and the path angle $\widehat{\phi}$ are closed to the target of them. When it comes to the convergence of the AoAs estimation network and the estimation of the AoA.

Proposition 2: Consider there is no interference between difference users. The estimation of the AoA must hold the proposition as follow:

$$\left[\left(\phi - \overline{\phi}\right) - \frac{2\pi(n-1)d}{\lambda}\left(\sin\theta - \sin\widehat{\theta}\right)\right] \rightarrow 2\pi l, l \in Z. \quad (8)$$

Proof: Please see our online appendix [11] for the complete proof. ■

Though the initialization process will ensure $\widehat{\theta} \rightarrow \theta$ at the beginning of the normal training process, the estimation of the path angle $\widehat{\phi}$ may not be accurate enough. The inaccuracy of the channel estimation network may impact the convergence of the AoAs estimation network during normal training process. That might cause AoAs estimation network to converge in the wrong direction which brings the multiplicity of global optima problem. Thus, the loss function (7) should be derived in more detail where $\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\phi}}$ are constant.

Proposition 3: Consider the single-user case (i.e., $K = 1$). There exist at least $N_{loc} = 2 \cdot \lfloor 2\frac{d}{\lambda} \rfloor + 1$ global optima that minimize the loss (7). More specifically, these K global optima have the same path gain and path angle but different AoAs as follows:

$$\theta^l = \arcsin\left(\sin\theta - \left\lceil -2\frac{d}{\lambda}\right\rceil + l\right), \ \forall l = 0, 1, ..., N_{loc} - 1. \quad (9)$$

Proof: Please see our online appendix [11] for the complete proof. ■

From the characterization of possible AoAs in Proposition 3, the number of global optima does not depend on the number of antenna elements, and increases with the carrier frequency. Therefore, increasing the number of antenna elements may not solve the problem, and the problem is more severe in millimeter wave systems of higher frequencies.

For the reason that there is no inter-users interference after initialization process, i.g. $\widehat{\theta} \rightarrow \theta$ is hold, the muti-users estimation fallback to the stacking of multiple single-user estimation problem. The number of multiple global optima can be regarded as $K \times N_{loc}$. To overcome this problem, this paper imposes the spatial filter in [12]. Thus, the sectorization method [13] is introduced into the proposed KA-AE network where the whole AoA estimation range is devided into Q sectors and the qth sector‘s range is $[-\pi/2 + (q-1)\pi/Q, -\pi/2 + q\pi/Q]$.

V. Numerical Simulation

This section verifies the theoretical analysis on the multiplicity of global optima, demonstrates the effectiveness of our proposed method in dealing with multiple global optima, and evaluate the performance against existing methods. In all simulations, we set $d/\lambda = 0.5$, snapshot $M = 40$ and the number of antennas $N_t = 32$, $Q = 3$, and $K = 3$ unless specified otherwise. We use 40000 samples for training and 10000 samples for testing.

A. Efficacy in Dealing With Multiplicity of Global Optima

Proposition 3 characterized the multiplicity of global optima and their positions in the AoA domain. Fig. 3 show the loss surface under different numbers of antennas, wavelengths, and antenna spacing. We can see that there are 3 and 9 distinct AoAs at global optima under $d/\lambda = 0.5$ and $d/\lambda = 2$, respectively, which is consistent with Proposition 3. We can also see that the multiplicity of global optima does not disappear as we increase the number of antennas, which validates our analysis.

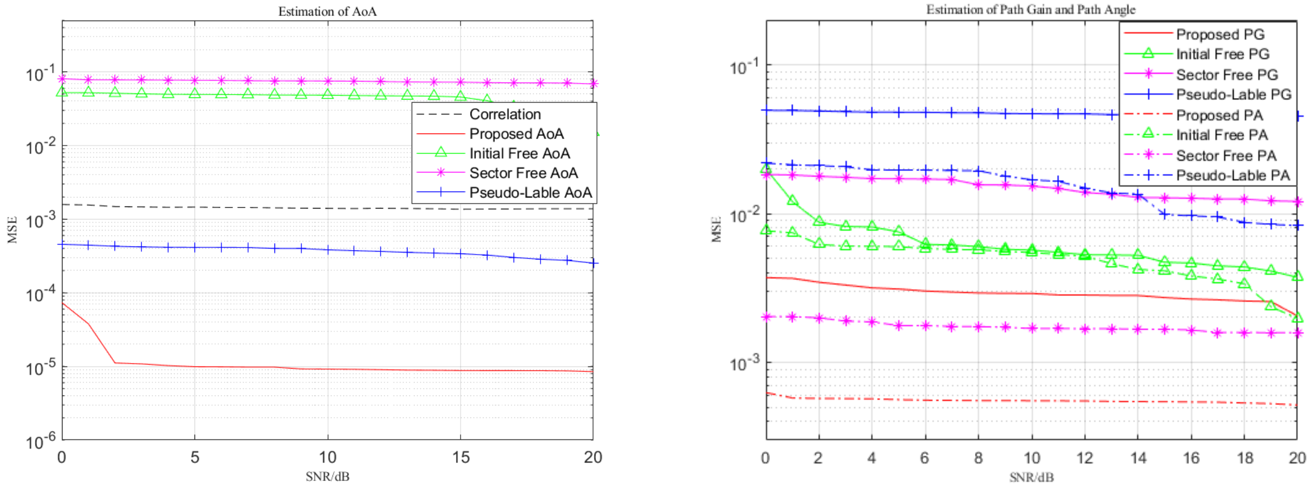
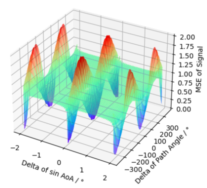
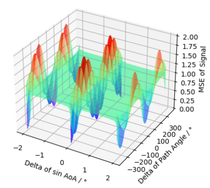


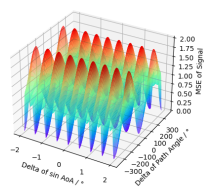
Fig. 2: Proposed is the performance of KA-AE and Initial Free means the network training without initialization process while Sector Free without sectorization. The Pseudo-Label uses (2) as loss function in the whole training process. Correlation means the AoA estimated by the correlation codebook in data preprocessing.



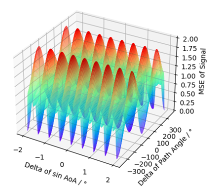
(a) 32 antennas with $d/\lambda = 0.5$



(b) 64 antennas with $d/\lambda = 0.5$



(c) 32 antennas with $d/\lambda = 2$



(d) 64 antennas with $d/\lambda = 2$

Fig. 3: Multiplicity of Global Optima.

Next, Fig. 2 shows that our proposed method effectively resolve the issue of global optima. In Fig. 2, "sector free" is the method without sectorization and has the highest estimation errors. Then we compare different algorithms under the sectorized scenario. The "Initial free" algorithm chooses the initial point randomly and achieves only negligible improvement over the "sector free" scenario. The "Pseudo-Label" method indicates the performance after the initialization phase, which further improves the performance. Our proposed method continues to minimize the training loss by alternating between improving the estimation of AoAs and path gains, and achieves significant performance improvement after the initialization.

B. Performance Improvement Over Existing Works

Finally, this paper compares the channel estimation accuracy of the KA-AE network with the following benchmarks:

- the MUSIC algorithm [14];
- a supervised learning algorithm, for which all labels of true AoAs, path angles, and path gains (called "All-Label Network") are used and the loss function is the mean squared error between the estimates and the labels.
- "AoA-Label Network", an variation of standard AE with labels of true AoAs, whose loss function is:

$$\mathbb{E}\Big[(\boldsymbol{Y}_R - \hat{\boldsymbol{Y}}_R)^2 + (\boldsymbol{Y}_I - \hat{\boldsymbol{Y}}_I)^2\Big] + \mathbb{E}\left[(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^2\right]. \quad (10)$$

The AoA-Label Network is used to demonstrate that the proposed KA-AE is able to enforce meaningful output of the encoder by the novel design of the decoder.

Fig. 4 summarizes the performance evaluation in terms of the signal reconstruction error and the estimation errors of AoAs, path gains, and path angles. Note that for the fully supervised All-Label Network, the signal reconstruction error is not shown because there is no decoder to reconstruct the signal. For the MUSIC algorithm, the reconstruction error and the estimation errors of path gains and path angles are not shown because the algorithm does not produce these estimates.

The key observation is that the proposed KA-AE achieves almost identical performance as the fully supervised All-Label Network. This demonstrates the advantage of the proposed method: the removal of labels in our method comes at almost no cost. This achievement is not trivial, because the classic unsupervised MUSIC algorithm has much higher AoA estimation errors.

The performance comparison with the AoA-Label Network illustrates that the proposed decoder enforces meaningful output of the encoder by incorporating the knowledge of the signal propagation model. The AoA-Label Network is the

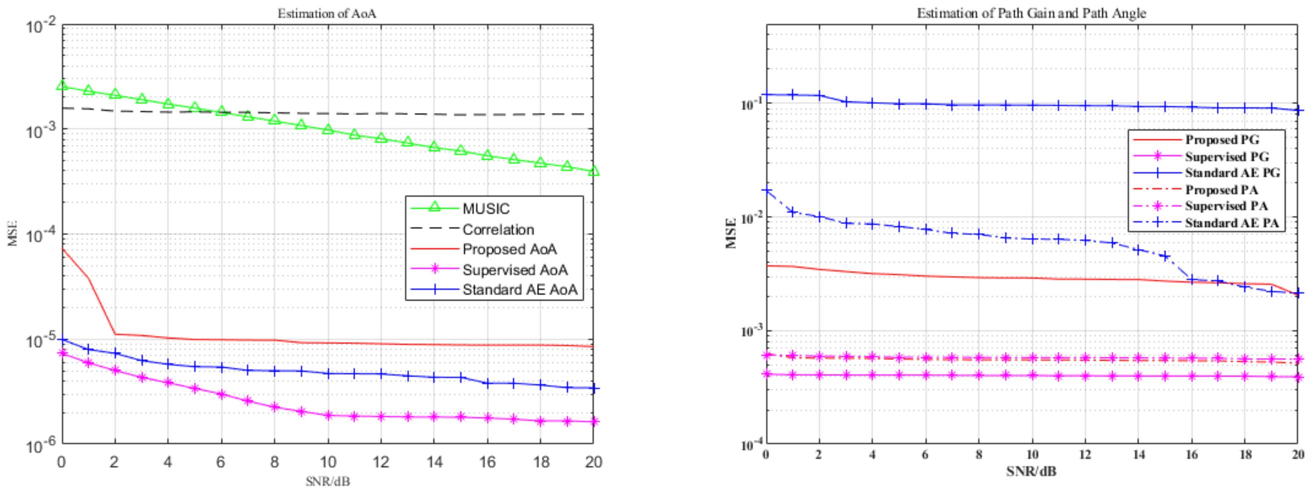


Fig. 4: Performance comparison.

standard AE with a modified loss function that includes the MSE of the AoA estimation. Note that in the AoA-Label Network, the decoder is the same as the one in the standard AE, which has learnable parameters. For standard AEs, the physical meanings of the embedding vector (i.e., the output of the encoder) cannot be controlled. For the AoA-Label Network, part of the embedding vector is enforced to be the AoA through the addition of the MSE of AoA estimation, but have no be controlled over the rest of the embedding vector. As expected and as demonstrated in Fig. 4, the AoA-Label Network has comparable performance with the proposed KA-AE in terms of AoA estimation, but has much worse performance in signal reconstruction and estimation of path gains and path angles.

VI. Conclusions

This paper proposed a novel variation of AE for unsupervised channel estimation in massive MIMO systems. The key design idea is to replace the decoder of the standard AE with the "known signal propagation model". Such knowledge is instilled in the proposed decoder by hardwiring it with the signal propagation model. Furthermore, correlation codebook is imposed to generate the pseudo-label and the training process is divided into initialization and normal training to eliminate the inter-users inference. Thereafter, the issue of multiple global optima is analyzed in the unsupervised estimation problem by separating the encoder into channel estimation network and AoAs estimation network as dual path network, and adopted the sectorization method to alleviate the issue. Numerical simulations were carried out to validate the theoretical analysis. Comparisons with carefully designed benchmark algorithms demonstrated that the proposed method achieves almost identical performance as the fully supervised method, and that the achievement is enabled by the preprocession of the data based on correlation, the mathematical analysis based on dual path encoder, the design of the training process and the novel design of the decoder.

References

[1] T. L. Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *IEEE Transactions on Wireless Communications*, 9(11):3590–3600, 2010.

[2] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. D. Doncker. The cost 2100 mimo channel model. *IEEE Wireless Communications*, 19(6):92–99, 2012.

[3] J. Li, Q. Zhang, X. Xin, Y. Tao, Q. Tian, F. Tian, D. Chen, Y. Shen, G. Cao, Z. Gao, and J. Qian. Deep learning-based massive mimo csi feedback. In *2019 18th International Conference on Optical Communications and Networks (ICOCN)*, pages 1–3, 2019.

[4] Ramina Ghods, Charles Jeon, Arian Maleki, and Christoph Studer. Optimal data detection and signal estimation in systems with input noise. *IEEE Transactions on Signal Processing*, 69:5105–5119, 2021.

[5] F. Roos, P. Hügler, L. L. T. Torres, C. Knill, J. Schlichenmaier, C. Vasanelli, and et al. Compressed sensing based single snapshot doa estimation for sparse mimo radar arrays. In *2019 12th German Microwave Conference (GeMiC)*, pages 75–78, March 2019.

[6] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh. Deep learning-based channel estimation. *IEEE Communications Letters*, 23(4):652–655, 2019.

[7] Rick Fritschek, Rafael F. Schaefer, and Gerhard Wunder. Reinforce security: A model-free approach towards secure wiretap coding. In *IEEE International Conference on Communications*, pages 1–6, 2021.

[8] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui. Deep learning for super-resolution channel estimation and doa estimation based massive mimo system. *IEEE Transactions on Vehicular Technology*, 67(9):8549–8560, 2018.

[9] Hao Ye, Le Liang, Geoffrey Ye Li, and Biing-Hwang Juang. Deep learning-based end-to-end wireless communication systems with conditional gans as unknown channels. *IEEE Transactions on Wireless Communications*, 19(5):3133–3143, 2020.

[10] Hengtao He, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Model-driven deep learning for mimo detection. *IEEE Transactions on Signal Processing*, 68:1702–1715, 2020.

[11] Zhiheng Guo, Yuanzhang Xiao, and Xiang Chen. Unsupervised channel estimation with dual path knowledge-aware auto-encoder. [Online Appendix]. https://yuanzhangxiao.github.io/Papers/ICCAppendix.pdf.

[12] B. D. Van Veen and K. M. Buckley. Beamforming: a versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(2):4–24, 1988.

[13] Q. He, L. Xiao, X. Zhong, and S. Zhou. Increasing the sum-throughput of cells with a sectorization method for massive mimo. *IEEE Communications Letters*, 18(10):1827–1830, 2014.

[14] P. Stoica and A. Nehorai. Music, maximum likelihood and cramer-rao bound: further results and comparisons. In *International Conference on Acoustics, Speech, and Signal Processing,*, pages 2605–2608 vol.4, 1989.