

Knowledge-Aided Deep Learning for Beam-space Channel Estimation in Millimeter-Wave Massive MIMO Systems

Xiuhong Wei, Chen Hu, and Linglong Dai

Abstract—Millimeter-wave massive multiple-input multiple-output (MIMO) can use a lens antenna array to considerably reduce the number of radio frequency (RF) chains, but channel estimation is challenging due to the number of RF chains is much smaller than the number of antennas. To address this challenge, we propose a beam-space channel estimation scheme based on deep learning (DL) in this paper. Specifically, the beam-space channel estimation problem can be formulated as a sparse signal recovery problem, which can be solved by the classical iterative algorithm named approximate message passing (AMP), and its corresponding version learned AMP (LAMP) realized by a deep neural network (DNN). Then, by exploiting the Gaussian mixture prior distribution of the beam-space channel elements, we derive a new shrinkage function to refine the classical AMP algorithm. Finally, by replacing the activation function in the conventional DNN with the derived Gaussian mixture shrinkage function, we propose a complex-valued Gaussian mixture LAMP (GM-LAMP) network specialized for estimating the beam-space channel. The simulation results show that, compared with the existing LAMP network and other conventional channel estimation schemes, the proposed GM-LAMP network considering the channel knowledge can improve the channel estimation accuracy with a low pilot overhead.

Index Terms—Millimeter-wave (mmWave), massive MIMO, beam-space channel estimation, approximate message passing (AMP), deep learning.

I. INTRODUCTION

Millimeter-wave (mmWave) massive multiple-input multiple-output (MIMO) has been considered as a key technique for 5G and beyond [1]. In order to reduce the hardware cost and power consumption caused by a large number of antennas and the associated radio frequency (RF) chains, the lens antenna array has been recently investigated to provide an energy-efficient realization of hybrid precoding for mmWave massive MIMO [2], [3]. By employing the lens antenna array, which can concentrate signals from different directions on different antennas, the spatial channel can be converted to the beam-space channel [4]. As there are only a few dominant propagation paths with large path gains at mmWave frequencies, the beam-space channel in mmWave massive MIMO systems is sparse in nature [5]. Therefore,

by only selecting a small number of dominant beams, the number of RF chains connected to the digital baseband can be considerably reduced. Beam selection requires the accurate channel state information in the beam-space [6], which is challenging due to the high channel dimension, especially when the number of RF chains is much smaller than the number of antennas [7]–[9].

A. Prior works

There are some recently proposed schemes for beam-space channel estimation. Specifically, [10] proposed a two-way channel estimation scheme with low computational complexity, where the antennas corresponding to the dominant beams are firstly determined by beam training between the base station (BS) and users, and then only channel elements corresponding to these selected antennas are estimated. However, the number of pilot symbols required to scan all possible beams is proportional to the number of BS antennas, which is very large (e.g., 256 antennas). Furthermore, by exploiting the sparsity of beam-space channels, some classical compressive sensing (CS) based schemes could estimate the beam-space channel with a reduced training overhead [11]–[13], such as the orthogonal matching pursuit (OMP) algorithm used in [11]. Apart from the sparsity, the beam-space channel may exhibit angular spreads. Based on this channel characteristics, [14] proposed a two-stage CS method for channel estimation, which consists of a matrix completion stage and a sparse recovery stage. What's more, by combining the structural characteristics and sparsity of the beam-space channel, a reliable support detection (SD)-based channel estimation scheme has been proposed in [15]. This scheme decomposes the total beam-space channel estimation problem into a series of subproblems, each of which only considers one channel path component.

Unfortunately, all of these beam-space channel estimation schemes above [11]–[15] cannot achieve the satisfying estimation accuracy in low signal-to-noise ratio (SNR) regions, and they also have high computational complexity when the dimension of the channel is high. As a powerful iterative algorithm for sparse signal recovery, the approximate message passing (AMP) algorithm can be used to recover the sparse high-dimensional beam-space channel with low computational complexity [16], [17]. However, it is difficult to find the optimal shrinkage parameters for the AMP algorithm (the empirical shrinkage parameters are usually used instead), which restricts its channel estimation performance in practice.

All authors are with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, P. R. China (e-mails: weixh19@mails.tsinghua.edu.cn, huc16@mails.tsinghua.edu.cn, daill@tsinghua.edu.cn).

This work was supported by the National Science and Technology Major Project of China under Grant 2018ZX03001004 and the National Natural Science Foundation of China for Outstanding Young Scholars under Grant 61722109.

Recently, the amazing success of deep learning (DL) in other fields like image recognition [18], [19] and speech processing [20] has greatly inspired researchers to use this powerful tool to solve some problems in wireless communications [21]–[24]. With the help of DL, we can extract underlying features of wireless big data and provide some improved solutions to some complicated problems in wireless communications, such as low density parity check (LDPC) decoding [21], sparse code multiple access (SCMA) codebook design [22], channel feedback for massive MIMO [23], and end-to-end communication [24].

Inspired by the powerful learning ability of deep neural network (DNN), [25] has presented a learned denoising-based approximate message passing (LDAMP) network for channel estimation, where a denoising convolutional neural network (DnCNN) for image recovery is incorporated into the AMP algorithm to replace the original shrinkage function. On the other hand, the empirical shrinkage parameters of the AMP algorithm can be directly optimized by DNN. Following this idea, the pioneering work [26] proposed a learned AMP (LAMP) network based on the classical AMP algorithm to jointly optimize its linear transform coefficients and nonlinear shrinkage parameters, so the improved sparse signal recovery performance can be achieved. The LAMP network is a direct transformation of the classical AMP algorithm by using the DNN, so it is general for any sparse signal recovery problems. However, for the specific beamspace channel estimation problem under investigation in this paper, the LAMP network cannot fully utilize the knowledge of the beamspace channel, and its beamspace channel estimation accuracy can be further improved.

B. Our contributions

In this paper, we study the beamspace channel estimation problem in mmWave massive MIMO systems by leveraging the DL tool, and propose a complex-valued knowledge-aided Gaussian mixture LAMP (GM-LAMP) network for improving the channel estimation performance¹. Firstly, the beamspace channel estimation can be formulated as a sparse signal recovery problem, which can be solved by the general AMP algorithm and the LAMP network. Then, by exploiting the knowledge that the beamspace channel elements follow the Gaussian mixture distribution, we derive a new shrinkage function, which is different from the original shrinkage function in the existing general LAMP network. Furthermore, integrating the LAMP network and the new shrinkage function derived from the Gaussian mixture distribution, we propose a knowledge-aided complex-valued GM-LAMP network to solve the beamspace channel estimation problem. By adopting the layer-by-layer training method, a large number of beamspace channels are used as the training data to optimize the GM-LAMP network. Finally, we provide the simulation results to show that compared with the conventional schemes, the proposed GM-LAMP network can achieve better channel estimation performance.

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

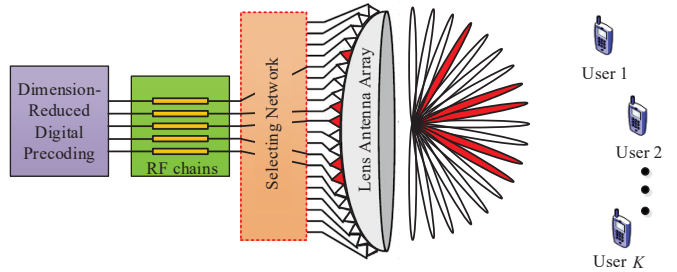


Fig. 1. mmWave massive MIMO with lens antenna array [15].

C. Organization and notation

The rest of the paper is organized as follows. Section II formulates the beamspace channel estimation problem in mmWave massive MIMO systems as a sparse signal recovery problem, and the conventional AMP algorithm and LAMP network for solving this problem will be reviewed. Based on the new shrinkage function derived from the Gaussian mixture distribution, the GM-LAMP network for improved beamspace channel estimation will be proposed in Section III. Simulation results will be provided to show the performance of the proposed GM-LAMP network in Section IV. Finally, conclusions are given in Section V.

Notation: Lower-case and upper-case boldface letters \mathbf{a} and \mathbf{A} denote a vector and a matrix, respectively; \mathbf{A}^H and \mathbf{A}^T denote the conjugate transpose and transpose of matrix \mathbf{A} , respectively; $\|\mathbf{a}\|_2$ denotes the l_2 -norm of vector \mathbf{a} ; $|a|$ denotes the amplitude of scalar a ; a^* denotes the conjugate of complex a ; $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of \mathbf{A} and \mathbf{B} ; $\mathcal{CN}(x; \mu, \sigma) \triangleq \frac{1}{\pi\sigma^2} e^{-\frac{(x-\mu)^2}{\sigma^2}}$ denotes the probability density function of the circularly symmetric complex Gaussian distribution with mean μ and variance σ^2 ; $\mathcal{U}(-a, a)$ denotes the probability density function of uniform distribution on $(-a, a)$; $\delta(x)$ denotes the Dirac delta function; $\text{sinc}(x) \triangleq \frac{\sin(N\pi x)}{N\pi x}$ denotes the Dirichlet sinc function. Finally, \mathbf{I}_K is the $K \times K$ identity matrix.

II. SYSTEM MODEL

In this section, we first introduce the beamspace channel model, and then formulate the beamspace channel estimation problem as a sparse signal recovery problem. Finally, the conventional AMP algorithm [16] and its corresponding LAMP network proposed in [26] to solve this problem will be introduced.

A. Beamspace channel

We consider a time division duplex (TDD) based mmWave massive MIMO system, as shown in Fig. 1 [15], where the BS employs a lens antenna array with N antennas and N_{RF} RF chains to simultaneously serve K single-antenna users.

In order to formulate the beamspace channel estimation problem, we start with the conventional mmWave massive MIMO channel in the spatial domain. According to the widely used Saleh-Valenzuela multipath channel model [11],

the channel vector \mathbf{h}_k of size $N \times 1$ between the k th ($k = 1, 2, \dots, K$) user and N -antenna BS can be presented by

$$\mathbf{h}_k = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \beta_{k,l} \mathbf{a}(\theta_{k,l}^{\text{azi}}, \theta_{k,l}^{\text{ele}}) = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \mathbf{c}_{k,l}, \quad (1)$$

where L_k is the number of resolvable paths, and $\mathbf{c}_{k,l} = \beta_{k,l} \mathbf{a}(\theta_{k,l}^{\text{azi}}, \theta_{k,l}^{\text{ele}})$ is the l th path component. $\beta_{k,l}$, $\theta_{k,l}^{\text{azi}}$ and $\theta_{k,l}^{\text{ele}}$ are the complex gain, azimuth and elevation of the l th path, respectively. $\mathbf{a}(\theta_{k,l}^{\text{azi}}, \theta_{k,l}^{\text{ele}})$ is the $N \times 1$ array steering vector, which depends on the array geometry. Ignoring the subscripts without loss of generality, for the simpler uniform linear arrays (ULAs), the array steering vector can be determined by one angle, which can be presented by [11]

$$\mathbf{a}_{\text{ULA}}(\theta) = \frac{1}{\sqrt{N}} \left[e^{-j2\pi d \sin(\theta) \mathbf{m} / \lambda} \right], \quad (2)$$

where $\mathbf{m} = [0, 1, \dots, N-1]^T$. For the widely considered uniform planar arrays (UPAs) with $N_1 \times N_2$ ($N = N_1 \times N_2$) antennas, we have [27]

$$\mathbf{a}_{\text{UPA}}(\theta^{\text{azi}}, \theta^{\text{ele}}) = \frac{1}{\sqrt{N}} \left[e^{-j2\pi d \sin(\theta^{\text{azi}}) \sin(\theta^{\text{ele}}) \mathbf{m}_1 / \lambda} \right] \otimes \left[e^{-j2\pi d \cos(\theta^{\text{ele}}) \mathbf{m}_2 / \lambda} \right]. \quad (3)$$

where $\mathbf{m}_1 = [0, 1, \dots, N_1-1]^T$ and $\mathbf{m}_2 = [0, 1, \dots, N_2-1]^T$. In (2) and (3), λ is the wavelength of carrier, and d is the antenna spacing usually satisfying $d = \lambda/2$ in mmWave communications [28]. Then, we can respectively define $\psi \triangleq d \sin(\theta) / \lambda$ as the spatial angle for ULAs, and $\psi^{\text{azi}} \triangleq d \sin(\theta^{\text{azi}}) \sin(\theta^{\text{ele}}) / \lambda$ and $\psi^{\text{ele}} \triangleq d \cos(\theta^{\text{ele}}) / \lambda$ as the spatial angles for UPAs [29].

The spatial domain channel can be directly transformed to the beamspace channel by using a lens antenna array. As a matter of fact, the lens antenna array plays the role of a spatial discrete fourier transform (DFT) matrix \mathbf{U} of size $N \times N$ [15]. For ULAs, the matrix \mathbf{U} can be expressed as

$$\mathbf{U} = [\bar{\mathbf{a}}_{\text{ULA}}(\bar{\psi}_1), \bar{\mathbf{a}}_{\text{ULA}}(\bar{\psi}_2), \dots, \bar{\mathbf{a}}_{\text{ULA}}(\bar{\psi}_N)]^H, \quad (4)$$

where $\bar{\psi}_n = \frac{1}{N} (n - \frac{N+1}{2})$ for $n = 1, 2, \dots, N$ are the spatial directions predefined by the lens antenna array. Similarly to (2), $\bar{\mathbf{a}}_{\text{ULA}}(\bar{\psi})$ can be presented by

$$\bar{\mathbf{a}}_{\text{ULA}}(\bar{\psi}) = \frac{1}{\sqrt{N}} \left[e^{-j2\pi \bar{\psi} \mathbf{m}} \right]. \quad (5)$$

For UPAs, \mathbf{U} can be expressed as

$$\mathbf{U} = [\bar{\mathbf{a}}_{\text{UPA}}(\bar{\psi}_1^{\text{azi}}, \bar{\psi}_1^{\text{ele}}), \dots, \bar{\mathbf{a}}_{\text{UPA}}(\bar{\psi}_{N_2}^{\text{azi}}, \bar{\psi}_{N_2}^{\text{ele}}), \dots, \bar{\mathbf{a}}_{\text{UPA}}(\bar{\psi}_{N_1}^{\text{azi}}, \bar{\psi}_1^{\text{ele}}), \dots, \bar{\mathbf{a}}_{\text{UPA}}(\bar{\psi}_{N_1}^{\text{azi}}, \bar{\psi}_{N_2}^{\text{ele}})]^H, \quad (6)$$

where $\bar{\psi}_n^{\text{azi}} = \frac{1}{N_1} (n - \frac{N_1+1}{2})$ for $n = 1, 2, \dots, N_1$ and $\bar{\psi}_n^{\text{ele}} = \frac{1}{N_2} (n - \frac{N_2+1}{2})$ for $n = 1, 2, \dots, N_2$ are respectively predefined spatial angles of azimuth and elevation by the lens antenna array. Similarly to (3), $\bar{\mathbf{a}}_{\text{UPA}}(\bar{\psi}^{\text{azi}}, \bar{\psi}^{\text{ele}})$ can be presented by

$$\bar{\mathbf{a}}_{\text{UPA}}(\bar{\psi}^{\text{azi}}, \bar{\psi}^{\text{ele}}) = \frac{1}{\sqrt{N}} \left[e^{-j2\pi \bar{\psi}^{\text{azi}} \mathbf{m}_1} \right] \otimes \left[e^{-j2\pi \bar{\psi}^{\text{ele}} \mathbf{m}_2} \right]. \quad (7)$$

Finally, the beamspace channel vector $\tilde{\mathbf{h}}_k$ of size $N \times 1$ between the k th user and the N -antenna BS can be presented by

$$\tilde{\mathbf{h}}_k = \mathbf{U} \mathbf{h}_k = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \tilde{\mathbf{c}}_{k,l}, \quad (8)$$

where $\tilde{\mathbf{c}}_{k,l} = \mathbf{U} \mathbf{c}_{k,l}$ is the l th channel component of beamspace channel.

B. Problem formulation

In order to acquire the channel state information, all users should transmit the known pilot symbols to the BS over Q instants. Due to the TDD channel reciprocity, we can only consider the uplink to formulate the channel estimation problem. Then, the downlink channel can be directly obtained according to the estimated uplink channel. In this paper, we adopt the widely used orthogonal pilot transmission strategy [15], where the uplink channel estimation for each user is independent due to the pilot orthogonality, and thus we can estimate the beamspace channel vectors between all K users and the BS one by one. Without loss of generality, we take the beamspace channel vector $\tilde{\mathbf{h}}_k$ between the k th user and the BS as an example to formulate the channel estimation problem.

In the q th instant for pilot transmission, the $N_{\text{RF}} \times 1$ measurement signal vector in the baseband at BS after beam selection can be presented as [15]

$$\mathbf{y}_{k,q} = \mathbf{A}_{k,q} \tilde{\mathbf{h}}_k s_{k,q} + \bar{\mathbf{n}}_{k,q}, \quad q = 1, 2, \dots, Q, \quad (9)$$

where $\mathbf{A}_{k,q}$ is the $N_{\text{RF}} \times N$ beam selection network, $s_{k,q}$ is the transmitted pilot symbol, $\bar{\mathbf{n}}_{k,q} = \mathbf{A}_{k,q} \mathbf{n}_{k,q}$ is the effective noise vector, where $\mathbf{n}_{k,q} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_N)$ is the $N \times 1$ noise vector with σ_n^2 representing the noise power.

After Q instants of pilot transmission, we can obtain the $M \times 1$ ($M = Q N_{\text{RF}}$) overall measurement vector by assuming $s_{k,q} = 1$ for $q = 1, 2, \dots, Q$ as

$$\mathbf{y}_k = \begin{bmatrix} \mathbf{y}_{k,1} \\ \mathbf{y}_{k,2} \\ \vdots \\ \mathbf{y}_{k,Q} \end{bmatrix} = \mathbf{A}_k \tilde{\mathbf{h}}_k + \mathbf{n}_k, \quad (10)$$

where $\mathbf{A}_k = [\mathbf{A}_{k,1}^T, \mathbf{A}_{k,2}^T, \dots, \mathbf{A}_{k,Q}^T]^T$ is the $M \times N$ selection matrix with the entry being $\pm \frac{1}{\sqrt{M}}$ [15], and $\mathbf{n}_k = [\bar{\mathbf{n}}_{k,1}^T, \bar{\mathbf{n}}_{k,2}^T, \dots, \bar{\mathbf{n}}_{k,Q}^T]^T$ is the $M \times 1$ effective noise vector in Q instants.

Since the channel estimation method is the same for all K users due to the pilot orthogonality, the subscript k in the problem (10) can be omitted, then (10) can be expressed as

$$\mathbf{y} = \mathbf{A} \tilde{\mathbf{h}} + \mathbf{n}. \quad (11)$$

Note that each element of the beamspace channel $\tilde{\mathbf{h}}$ in (11) can be regarded as the sum of all paths at each spatial angle. As there are only a few dominant propagation paths with large gains due to limited scattering at mmWave frequencies [5], the beamspace channel $\tilde{\mathbf{h}}$ is sparse. Consequently, we can apply the sparse signal recovery algorithms in CS to estimate the beamspace channel with a low pilot overhead, where the

matrix \mathbf{A} in (11) can be regarded as the sensing matrix in CS. That is to say, the beamspace channel estimation problem in (11) can be formulated as a sparse signal recovery problem

$$\min \|\tilde{\mathbf{h}}\|_0, \quad \text{s.t.} \|\mathbf{y} - \mathbf{A}\tilde{\mathbf{h}}\|_2 \leq \varepsilon, \quad (12)$$

where $\|\tilde{\mathbf{h}}\|_0$ is the number of non-zero elements of $\tilde{\mathbf{h}}$, ε is the error tolerance parameter.

C. AMP algorithm and LAMP network

Since the number of antennas in mmWave massive MIMO systems is usually large, the dimension of sparse signal in (12) is high. As an iterative algorithm in CS, AMP can be used to recover the sparse signals, especially for high-dimensional sparse signals. In this subsection, we introduce how the complex-valued AMP algorithm estimates the beamspace channel, as shown in **Algorithm 1**.

Algorithm 1: Approximate Message Passing (AMP)

Input: The measurement vector \mathbf{y} , the sensing matrix \mathbf{A} , the number of iterations $\hat{z}T$.

Initialization: $\mathbf{v}_0 = \mathbf{0}$, $b_0 = 0$, $\hat{\mathbf{h}}_0 = \mathbf{0}$.

for $t = 1, 2, \dots, \hat{z}T$ **do**

1. $\mathbf{v}_t = \mathbf{y} - \mathbf{A}\hat{\mathbf{h}}_{t-1} + b_{t-1}\mathbf{v}_{t-1}$

2. $\sigma_t^2 = \frac{1}{\hat{z}M} \|\mathbf{v}_t\|_2^2$

3. $\mathbf{r}_t = \hat{\mathbf{h}}_{t-1} + \mathbf{A}^T \mathbf{v}_t$

4. $\hat{\mathbf{h}}_t = \boldsymbol{\eta}_{\text{st}}(\mathbf{r}_t; \lambda_t, \sigma_t^2)$

5. $b_t = \frac{N}{M} \left\langle \boldsymbol{\eta}'_{\text{st}}(\mathbf{r}_t; \lambda_t, \sigma_t^2) \right\rangle$

end for

Output: Sparse signal recovery results: $\hat{\mathbf{h}} = \hat{\mathbf{h}}_T$.

In **Algorithm 1**, the term $b_{t-1}\mathbf{v}_{t-1}$ in step 1 is called Onsager Correction [16], which is introduced into the AMP algorithm to accelerate the convergence. The critical step of the AMP algorithm is step 4, in which the estimate $\hat{\mathbf{h}}_t$ in the t th iteration is obtained through the soft threshold shrinkage function $\boldsymbol{\eta}_{\text{st}}: \mathbb{C}^N \rightarrow \mathbb{C}^N$. The shrinkage function $\boldsymbol{\eta}_{\text{st}}$ is nonlinear element-wise operation, which takes the sparsity of the vector $\tilde{\mathbf{h}}$ into consideration, and makes the estimate $\hat{\mathbf{h}}_t$ sparser. For the i th element $r_{t,i} = |r_{t,i}| e^{j\omega_{t,i}}$ ($i = 1, 2, \dots, N$) of input vector \mathbf{r}_t , we have:

$$\begin{aligned} & [\boldsymbol{\eta}_{\text{st}}(\mathbf{r}_t; \lambda_t, \sigma_t^2)]_i \\ &= \boldsymbol{\eta}_{\text{st}}(|r_{t,i}| e^{j\omega_{t,i}}; \lambda_t, \sigma_t^2) \\ &= \max(|r_{t,i}| - \lambda_t \sigma_t, 0) e^{j\omega_{t,i}}, \end{aligned} \quad (13)$$

where $\omega_{t,i}$ is the phase of complex-valued element $r_{t,i}$, λ_t is the predefined and fixed parameter in the t th iteration, and σ_t^2 is updated via estimating the noise variance in step 2. In step 5, $\boldsymbol{\eta}'_{\text{st}}$ is the element-wise derivative of $\boldsymbol{\eta}_{\text{st}}$, which can be presented by

$$\left[\boldsymbol{\eta}'_{\text{st}}(\mathbf{r}_t; \lambda_t, \sigma_t^2) \right]_i = \frac{\partial [\boldsymbol{\eta}_{\text{st}}(\mathbf{r}_t; \lambda_t, \sigma_t^2)]_i}{\partial r_{t,i}}, \quad (14)$$

and $\langle \cdot \rangle$ denotes the average value of all elements of the vector. From (13), we can find that the soft threshold shrinkage

function can shrink the amplitude of complex-valued input with low power to zero.

Although the AMP algorithm is good at dealing with large-scale sparse signal recovery problem, there are still two problems when it is used for sparse beamspace channel estimation. First, the shrinkage parameter λ_t in (13) usually takes same empirical values for all iterations, which limits the performance of the AMP algorithm. Second, the general AMP algorithm cannot exploit the inherent knowledge in the specific beamspace channel estimation problem (12).

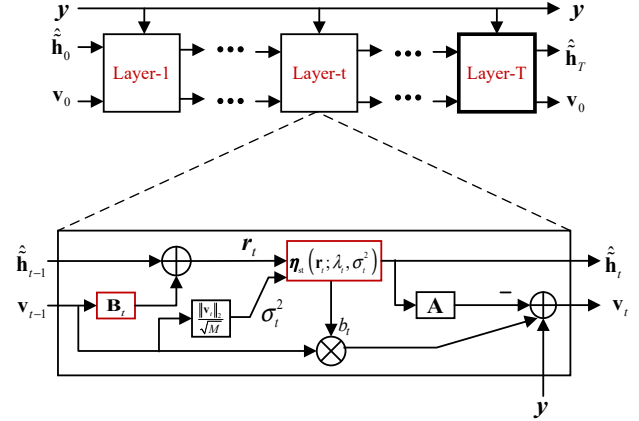


Fig. 2. LAMP network structure (the t th layer is explained in detail) [26].

To solve the first problem, the LAMP network based on the classical AMP algorithm has been recently proposed to optimize the nonlinear shrinkage parameter λ_t in each iteration [26]. As shown in Fig. 2, each iteration of the classical AMP algorithm is mapped to each layer of the LAMP network. To be specific, the inputs of the t th layer are $\mathbf{y} \in \mathbb{C}^M$, $\hat{\mathbf{h}}_{t-1} \in \mathbb{C}^N$ and $\mathbf{v}_{t-1} \in \mathbb{C}^M$, where \mathbf{y} is the measurement vector in (11), $\hat{\mathbf{h}}_{t-1}$ and \mathbf{v}_{t-1} are the outputs of the previous $(t-1)$ th layer. Following the principle of the AMP algorithm, each layer of the LAMP network processes the signal as follows, which is similar to **Algorithm 1**:

$$\mathbf{v}_t = \mathbf{y} - \mathbf{A}\hat{\mathbf{h}}_{t-1} + b_{t-1}\mathbf{v}_{t-1}, \quad (15)$$

$$\sigma_t^2 = \frac{1}{M} \|\mathbf{v}_t\|_2^2, \quad (16)$$

$$\mathbf{r}_t = \hat{\mathbf{h}}_{t-1} + \mathbf{B}_t \mathbf{v}_t, \quad (17)$$

$$\hat{\mathbf{h}}_t = \boldsymbol{\eta}_{\text{st}}(\mathbf{r}_t; \lambda_t, \sigma_t^2), \quad (18)$$

$$b_t = \frac{N}{M} \left\langle \boldsymbol{\eta}'_{\text{st}}(\mathbf{r}_t; \lambda_t, \sigma_t^2) \right\rangle, \quad (19)$$

where the shrinkage function $\boldsymbol{\eta}_{\text{st}}$ of the AMP algorithm plays a role of the nonlinear activation function in the conventional DNN [26]. What's more, from (17), we can find that different from the step 3 in **Algorithm 1**, the LAMP network can choose the different coefficients \mathbf{B}_t for each layer t , which can replace \mathbf{A}^T as the linear transform from the measurement signal space to the original sparse signal space. It is worth noting that \mathbf{A}^T is selected only for the convenience of derivation in

AMP algorithm. In the training stage of the LAMP network, the linear transform coefficients \mathbf{B}_t of size $N \times M$ in (17) and the nonlinear shrinkage parameters λ_t in (18) and (19) can be optimized. Therefore, given the enough training data, the LAMP network can find the best shrinkage parameters by leveraging the powerful learning ability of DNN.

However, the second problem of the AMP algorithm for channel estimation has not been solved. The conventional AMP algorithm and its corresponding LAMP network only consider the sparsity of the signals to be recovered, which are general for any sparse signal recovery problem. In particular, compared with the activation function without an explicit physical meaning in conventional DNN, the shrinkage function of the LAMP network is not specifically designed for the beamspace channel estimation problem under investigation. In order to solve the problem, we will utilize the prior knowledge of sparse beamspace channel to propose a more suitable network for the beamspace channel estimation in mmWave massive MIMO systems in the next section.

III. PROPOSED GM-LAMP NETWORK FOR BEAMSPACE CHANNEL ESTIMATION

In this section, we will propose a knowledge-aided GM-LAMP network for beamspace channel estimation by exploiting the beamspace channel elements' prior distribution. First, a new shrinkage function will be derived according to the Gaussian mixture distribution of the beamspace channel. Second, the offline training phase and online estimation phase of the proposed GM-LAMP network will be discussed in detail. Finally, we will also discuss how to extend the idea of GM-LAMP network for other sparse signal recovery problems.

A. Gaussian mixture distribution and its corresponding shrinkage function

As we all know, we are likely to get a more accurate estimate with more prior knowledge of the channel. Next, we will utilize more specific prior distribution (besides sparsity) of beamspace channel to refine the LAMP network.

There have been some previous works to consider the Gaussian mixture distribution to model the prior distribution of beamspace channel elements for ULAs [13] and for UPAs [30] and verify its validity. Specifically, the probability density function of the element \tilde{h} of the beamspace channel $\tilde{\mathbf{h}}$ can be expressed as:

$$p(\tilde{h}; \boldsymbol{\theta}) = \sum_{k=0}^{N_c-1} p_k \mathcal{CN}(\tilde{h}; 0, \sigma_k^2), \quad (20)$$

where $\boldsymbol{\theta} = \{p_0, p_1, \dots, p_{N_c-1}, \sigma_0^2, \sigma_1^2, \dots, \sigma_{N_c-1}^2\}$ is the set of all distribution parameters. N_c is the number of Gaussian components in the Gaussian mixture distribution, p_k is the probability of k th Gaussian component, and σ_k^2 denotes the variance of the k th Gaussian component. Take the ULA as an example, the rationality of the Gaussian mixture distribution can be explained based on the following two observations.

From (1), (4) and (8), the n th element \tilde{h}_n of the beamspace channel $\tilde{\mathbf{h}}$ can be expressed by

$$\tilde{h}_n = \sqrt{\frac{N}{L}} \sum_{l=1}^L \beta_l \text{sinc}(\Delta\psi_n), \quad (21)$$

where $\Delta\psi_n = \bar{\psi}_n - \psi_l$. Firstly, it is noted that the complex gain β_l follows the zero-mean complex Gaussian distribution. Secondly, when the practical spatial direction ψ_l for the l th path is close to the predefined spatial direction $\bar{\psi}_n$, $\text{sinc}(\Delta\psi_n)$ has a large value, which brings the large power for \tilde{h}_n . Similarly, when the practical spatial direction ψ_l for the l th path is far away from the predefined spatial direction $\bar{\psi}_n$, $\text{sinc}(\Delta\psi_n)$ has a small value, which brings the small power for \tilde{h}_n . It is due to the random of the practical spatial direction ψ_l that the different \tilde{h}_n can be regarded as the different Gaussian component with different variance. So, the Gaussian mixture distribution is expected to model the distribution of the beamspace channel elements.

It is worth noting that when the variance of a Gaussian component is zero, the probability density function of Gaussian distribution will be changed to

$$\mathcal{CN}(\tilde{h}; 0, 0) = \delta(\tilde{h}), \quad (22)$$

where the $\delta(\tilde{h})$ is the Dirac delta function, which means the random variable \tilde{h} will be exact zero. Thus, the Gaussian mixture distribution can also describe the sparsity of the beamspace channel as a special case.

Then, we can derive the scalar version $\eta_{\text{gm}}: \mathbb{C} \rightarrow \mathbb{C}$ of element-wise Gaussian mixture shrinkage function based on the the Bayesian minimize mean square error (MMSE) estimation principle [16] as follows:

$$\eta_{\text{gm}} = \mathbb{E}\{\tilde{h} | r; \boldsymbol{\theta}, \sigma^2\} = \frac{\int \tilde{h} p(r | \tilde{h}; \sigma^2) p(\tilde{h}; \boldsymbol{\theta}) d\tilde{h}}{\int p(r | \tilde{h}; \sigma^2) p(\tilde{h}; \boldsymbol{\theta}) d\tilde{h}}, \quad (23)$$

where the input element r of the shrinkage function is modeled by [26]

$$r = \tilde{h} + n, \quad (24)$$

where n is the additive Gaussian noise following $\mathcal{CN}(0, \sigma^2)$. Thus, we have

$$p(r | \tilde{h}; \sigma^2) = \mathcal{CN}(r; \tilde{h}, \sigma^2). \quad (25)$$

Given $p(\tilde{h}; \boldsymbol{\theta})$ by (20), we have

$$\begin{aligned} & p(r | \tilde{h}; \sigma^2) p(\tilde{h}; \boldsymbol{\theta}) \\ &= \mathcal{CN}(r; \tilde{h}, \sigma^2) \sum_{k=0}^{N_c-1} p_k \mathcal{CN}(\tilde{h}; 0, \sigma_k^2) \\ &= \sum_{k=0}^{N_c-1} p_k \mathcal{CN}(r; \tilde{h}, \sigma^2) \mathcal{CN}(\tilde{h}; 0, \sigma_k^2) \\ &= \sum_{k=0}^{N_c-1} p_k \mathcal{CN}(r; 0, \sigma^2 + \sigma_k^2) \mathcal{CN}(\tilde{h}; \tilde{\mu}_k(r), \tilde{\sigma}_k^2), \end{aligned} \quad (26)$$

where $\tilde{\mu}_k(r) = \frac{\sigma_k^2 r}{\sigma^2 + \sigma_k^2}$ and $\tilde{\sigma}_k^2(r) = \frac{\sigma^2 \sigma_k^2}{\sigma^2 + \sigma_k^2}$.

Finally, by substituting (26) in (23), we can derive a new shrinkage function based on the Gaussian mixture distribution as:

$$\eta_{\text{gm}}(r; \boldsymbol{\theta}, \sigma^2) = \frac{\sum_{k=0}^{N_c-1} p_k \tilde{\mu}_k(r) \mathcal{CN}(r; 0, \sigma^2 + \sigma_k^2)}{\sum_{k=0}^{N_c-1} p_k \mathcal{CN}(r; 0, \sigma^2 + \sigma_k^2)}, \quad (27)$$

where a set of all distribution parameters $\boldsymbol{\theta}$ can also be called as the shrinkage parameters. Compared with the general soft threshold shrinkage function η_{st} in the existing LAMP network, the Gaussian mixture shrinkage function η_{gm} considering the prior knowledge of the beamspace channel is designed for the specific beamspace channel estimation problem.

Now we have derived the Gaussian mixture shrinkage function, based on which we will propose the GM-LAMP network for the beamspace channel estimation in the next subsection.

B. Proposed GM-LAMP network

In order to estimate beamspace channel more accurately, we integrate the LAMP network and the new shrinkage function derived from the Gaussian mixture distribution to propose a knowledge-aided GM-LAMP network. Specifically, we replace the original soft threshold shrinkage function in the existing LAMP network by the Gaussian mixture shrinkage function. Therefore, the proposed GM-LAMP network is still constructed on the AMP algorithm. But it utilizes the Gaussian distribution prior knowledge of beamspace channel by leveraging the Gaussian mixture shrinkage function, which can improve channel estimation performance.

Next, we discuss how the GM-LAMP network works for the beamspace channel estimation in mmWave massive MIMO systems. The GM-LAMP network mainly works in two phases: offline training phase and online estimation phase. In the offline training phase, given a large number of known training data, the GM-LAMP network aims to learn the linear transform coefficients \mathbf{B}_t and nonlinear shrinkage parameters $\boldsymbol{\theta}_t$ by minimizing the loss function. In the estimation phase, by inputting the new measurements, the trained GM-LAMP network can output the estimated beamspace channel. Next, we introduce these two phases in detail.

1) *Offline training phase:* In this paper, we adopt the supervised learning to train the GM-LAMP network. Specifically, we first generate the training data set $\{\mathbf{y}^d, \hat{\mathbf{h}}^d\}_{d=1}^D$ according to (1) and (11), where \mathbf{y}^d is the input of the GM-LAMP network, and $\hat{\mathbf{h}}^d$ is the corresponding label. Since the GM-LAMP network is constructed based on the AMP algorithm, the layer-by-layer training is used to jointly optimize the linear transform coefficients \mathbf{B}_t and nonlinear shrinkage parameters $\boldsymbol{\theta}_t$ [26]. Each layer of the GM-LAMP network includes a linear transform operation and nonlinear shrinkage operation. Thus, the overall trainable variables set is $\Omega_T = \{\mathbf{B}_t, \boldsymbol{\theta}_t\}_{t=1}^T$, where \mathbf{B}_t and $\boldsymbol{\theta}_t$ are the trainable variables of the t th layer. Different from the general DNN, which only defines one loss function for the whole network, each layer in GM-LAMP network defines two loss functions, which are related

to the linear transform operation and the nonlinear shrinkage operation separately. The linear loss function and the nonlinear loss function of the t th layer can be defined respectively as

$$L_t^{\text{linear}}(\Omega_t) = \frac{1}{D} \sum_{d=1}^D \left\| \mathbf{r}_t^d(\mathbf{y}^d, \Omega_t) - \hat{\mathbf{h}}^d \right\|, \quad (28)$$

$$L_t^{\text{nonlinear}}(\Omega_t) = \frac{1}{D} \sum_{d=1}^D \left\| \hat{\mathbf{h}}_t^d(\mathbf{y}^d, \Omega_t) - \hat{\mathbf{h}}^d \right\|, \quad (29)$$

where \mathbf{r}_t^d is the input of the shrinkage function after the linear transform operation, and $\hat{\mathbf{h}}_t^d$ is the output of the shrinkage function (i.e., the estimated channel of the t th layer). It is note that the loss function of the t th layer has nothing to do with the trainable variables from the $(t+1)$ th layer to T th layer.

Algorithm 2: Layer-by-Layer Training Method

Initialization: $\mathbf{B}_1 = \mathbf{A}^T, \boldsymbol{\theta}_1 = \boldsymbol{\theta}^0$.

1. Learn \mathbf{B}_1 to minimize L_1^{linear}
 2. Learn $\boldsymbol{\theta}_1$ with fixed \mathbf{B}_1 to minimize $L_1^{\text{nonlinear}}$
 3. Re-learn Ω_1 to minimize $L_1^{\text{nonlinear}}$
- for** $t = 2, \dots, T$ **do**
4. Initialization: $\mathbf{B}_t = \mathbf{B}_{t-1}, \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$
 5. Learn \mathbf{B}_t with fixed Ω_{t-1} to minimize L_t^{linear}
 6. Re-learn $\{\Omega_{t-1}, \mathbf{B}_t\}$ to minimize L_t^{linear}
 7. Learn $\boldsymbol{\theta}_t$ with fixed $\Omega_{t-1}, \mathbf{B}_t$ to minimize $L_t^{\text{nonlinear}}$
 8. Re-learn Ω_t to minimize $L_t^{\text{nonlinear}}$

end for

Output: Ω_T .

Algorithm 2 shows the specific layer-by-layer training method. The training of each layer t includes a linear training based on minimizing the linear loss function L_t^{linear} and a nonlinear training based on minimizing the nonlinear loss function $L_t^{\text{nonlinear}}$, where the linear transform coefficients \mathbf{B}_t or nonlinear shrinkage parameters $\boldsymbol{\theta}_t$ are first optimized individually, and then all the trainable variables of the previous t layers are optimized globally. Step 5 represents the individual optimization of the linear transform coefficients \mathbf{B}_t , where only the \mathbf{B}_t is trainable with the trainable variables Ω_{t-1} of previous $(t-1)$ layers unchanged. Step 6 represents the global optimization for the linear training, where the trainable variables of previous t layers are re-learned. Similarly, step 7 and step 8 represent the individual optimization and the global optimization for the nonlinear training, respectively.

After all trainable variables Ω_T of T layers are optimized, we can obtain a trained GM-LAMP network to directly estimate the beamspace channel.

2) *Online estimation phase:* In this phase, we apply the trained GM-LAMP network to the beamspace channel estimation problem in mmWave massive MIMO systems. According to (11), we can obtain new measurements different from the training data set for different users. Then, the new measurements for different users are fed into the GM-LAMP network in turn to directly generate the beamspace channel estimate.

Finally, the normalized mean square error (NMSE) is used to evaluate the performance of the GM-LAMP network:

$$\text{NMSE} = \mathbb{E} \left\{ \frac{\sum_{k=1}^K \|\hat{\mathbf{h}}_k - \tilde{\mathbf{h}}_k\|_2^2}{\sum_{k=1}^K \|\tilde{\mathbf{h}}_k\|_2^2} \right\}. \quad (30)$$

C. Insights from the proposed GM-LAMP network

From the discussion above, we can find that in the existing LAMP network [26], the soft threshold shrinkage function only utilizes the sparsity of the signal to be recovered. In contrast, the Gaussian mixture shrinkage function in the proposed GM-LAMP network is derived from the Gaussian mixture distribution, which can approximate the distribution of beamspace channel elements more accurately. With the help of more channel knowledge, the GM-LAMP network is more suitable for the beamspace channel estimation.

In this paper, we refine the existing LAMP network based on prior knowledge of the beamspace channel to improve the estimation accuracy. This idea can be extended to solve other sparse signal recovery problems in wireless communications with improved performance. If we know the prior distribution of sparse signals, e.g., the sparse active users in massive machine-type communications, the sparse active antennas in spatial modulation systems, and the sparse interfering BSs in ultra-dense networks [31], we can obtain a new shrinkage function based on the new distribution for the DNN, thus the performance can be improved.

Moreover, the most existing DNNs, such as the fully connected network, have a generality for a large number of problems, but are not optimized for the specific problems to be solved. For the specific problem, by leveraging the domain knowledge (besides the signal distribution considered in this paper, e.g., other statistics like mean and variance, the inherent correlation of the signal, etc.), we can design some specialized DNNs for the specific problems with better performance.

IV. SIMULATION RESULTS

In this section, we present the beamspace channel estimation performance comparison among the proposed GM-LAMP network, the existing LAMP network, and other conventional channel estimation schemes.

A. Simulation setup

In our simulations, we consider that the BS equips a $N = 256$ lens antenna array and $N_{\text{RF}} = 16$ RF chains. The number of the single-antenna users is set to $K = 16$. For the k th user, we generate the spatial channel as follows [7]: 1) $L_k = 3$ path components; 2) $\beta_{k,l} \sim \mathcal{CN}(0, 1)$ for $l = 1, 2, 3$; 3) $\theta_{k,l} \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$, $\theta_{k,l}^{\text{azi}} \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$ and $\theta_{k,l}^{\text{ele}} \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$ for $l = 1, 2, 3$. Finally, the SNR are defined as $1/\sigma_n^2$.

In this paper, the GM-LAMP network is composed of $T = 8$ layers, where each layer has the same connected method similar to the iterative process of the AMP algorithm. The

number of nodes for every layer is depended on the number of the measurements M and the dimension of the beamspace channel N , i.e., $M + N$. The number of Gaussian components in the Gaussian mixture shrinkage function η_{gm} is set to 4. In **Algorithm 2**, the initialization θ^0 of the trainable variables θ_1 consists of 8 parameters representing the probability and variance of four Gaussian components, where we set one of the variances to 0 considering the sparsity of the beamspace channel and other initializations are randomly selected.

The GM-LAMP network are implemented with Tensorflow using a P100 GUP. The training, validation, and testing sets contain 80000, 2000, 2000 samples, respectively. The method of training is layer-by-layer, as mentioned in the Section III-B. In the offline training phase, we adopt the Adam optimizer [26] to optimize the trainable variables, and use mini-batches of 128 samples for each updating. The training rate for individual optimization is set at 0.001, and for the global optimization, the training rate decreases to 0.0005, 0.0001 and 0.00001 in turn when the validation error is no longer reduced. According to our experiments, a complete training takes about 9 hours under the above configure.

For the OMP-based channel estimation scheme, in order to determine the number of its iterations, we assume that the sparsity level of the beamspace channel is 24. For the SD-based channel estimation scheme, we follow [15] to retain 8 strongest elements for each beamspace channel component, thus the sparsity level is also 24 for the complete beamspace channel. For the AMP-based channel estimation scheme, the shrinkage parameters $\lambda_t = 1.1402$ for each layer t , as did in [26]. The configure of the LAMP network is the same as that of the GM-LAMP network. For the specific sensing matrix \mathbf{A} , we trained two LAMP networks and two GM-LAMP networks to work at 0 ~ 10 dB SNR regions and 10 ~ 20 dB SNR regions, respectively.

B. Simulation results

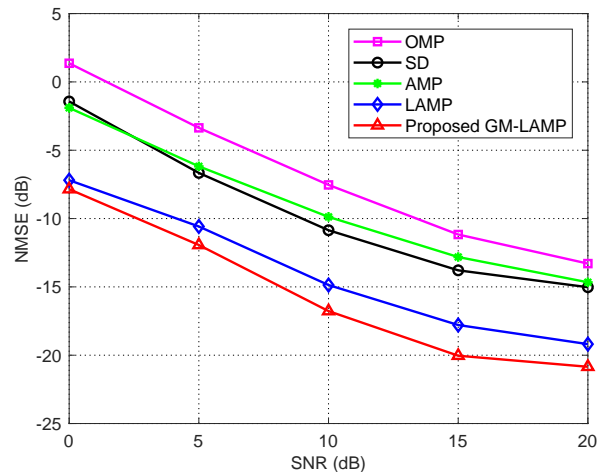


Fig. 3. NMSE performance comparison against different SNR for ULAs.

First, we compare the proposed GM-LAMP network with the OMP algorithm [11], the SD algorithm [15], the AMP algorithm [16], and the existing LAMP network [26], where

the ULA is considered. Fig. 3 shows the NMSE performance comparison of four different beamspace channel estimation schemes mentioned above against the SNR. The pilot transmission instants are all set to $Q = 8$, i.e., the length of measurement vector \mathbf{y} is $M = QN_{RF} = 128$, which is only the half of the length of the beamspace channel vector \mathbf{h} . From Fig. 3, we can observe that compared with all existing schemes under investigation, the proposed GM-LAMP network enjoys lower estimation error in all considered SNR regions. In particular, we can observe that the NMSE performance of the OMP-based channel estimation scheme, the AMP-based channel estimation scheme and the SD-based channel estimation scheme is poor, whereas the two DL-based channel estimation schemes LAMP and GM-LAMP can achieve better NMSE performance in low SNR regions (e.g., less than 10 dB). Moreover, due to the consideration of prior knowledge of the beamspace channel, the knowledge-aided GM-LAMP network has better channel estimation accuracy than the LAMP network. For example, when the SNR is 15 dB, the NMSE performance achieved by the LAMP network is -17.8 dB, while the NMSE performance achieved by the proposed GM-LAMP network is -20 dB.

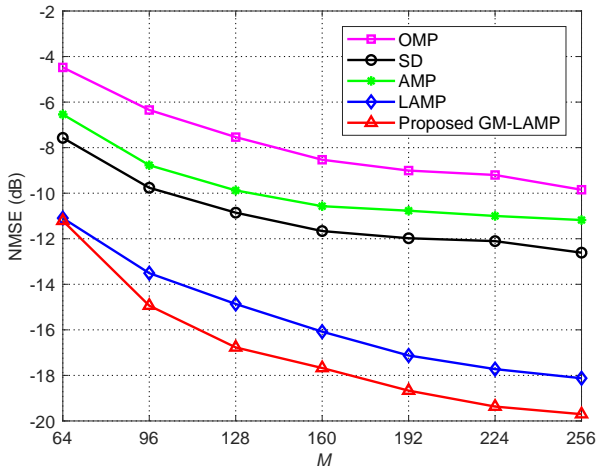


Fig. 4. NMSE performance comparison against different number of measurements M for ULAs.

Fig. 4 shows the NMSE performance comparison of five different beamspace channel estimation schemes against the number of measurements M , where the SNR is set to 10 dB. Given the number of RF chains N_{RF} , the number of measurements M is related to the number of pilot transmission instants Q ($M = QN_{RF}$), which means that the smaller M leads to smaller the pilot transmission instants Q . According to Fig. 4, we can find that to obtain the same beamspace channel estimation NMSE performance, the number of measurements M required by the proposed GM-LAMP network is smaller than the OMP, AMP and the SD algorithm and the LAMP network. For example, the NMSE performance achieved by the GM-LAMP network with $M = 96$ ($Q = 6$) is better than that of achieved by the conventional OMP, AMP and SD algorithm with $M = 256$ ($Q = 16$). Besides, owing to the use of the Gaussian mixture shrinkage function, the proposed GM-LAMP network obviously outperforms the existing LAMP

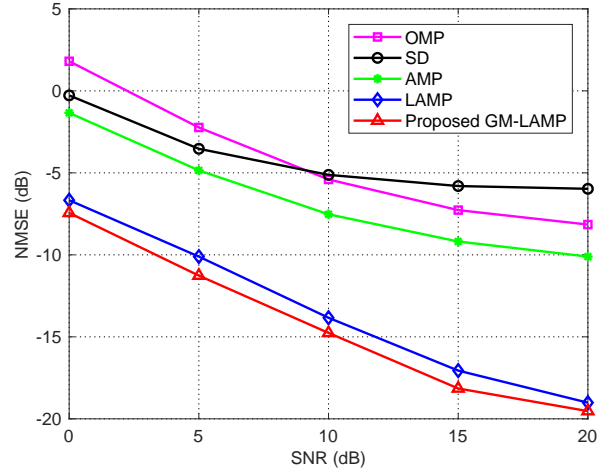


Fig. 5. NMSE performance comparison against different SNR for UPAs.

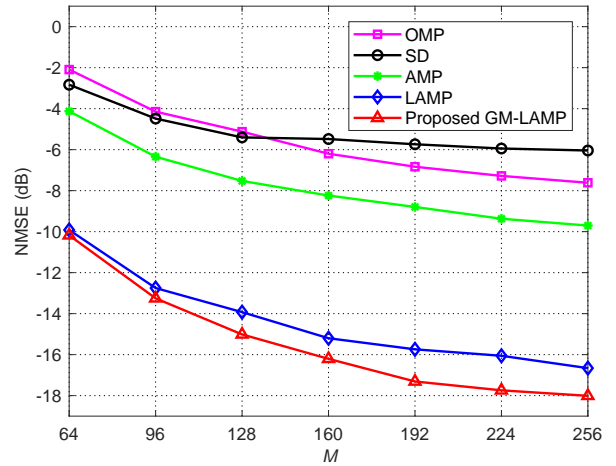


Fig. 6. NMSE performance comparison against different number of measurements M for UPAs.

network. For example, the NMSE performance achieved by the GM-LAMP network with $M = 128$ ($Q = 8$) is almost the same as that of achieved by the LAMP network with $M = 192$ ($Q = 12$). Therefore, we can conclude that the proposed GM-LAMP network can achieve the satisfying channel estimation performance with lower pilot overhead compared with other channel estimation schemes.

In Figs. 5-6, we compare the NMSE performance of five different beamspace channel estimation schemes when the 16×16 UPA is considered. The number of the measurements M in Fig. 5 and the SNR in Fig. 6 are fixed to 128 and 10 dB, respectively. We can find the conventional OMP, AMP and SD algorithms cannot achieve the satisfying beamspace channel estimation performance when the antenna array is UPA. Especially, the SD algorithm only proposed for ULAs performs even worse than the OMP algorithm. In contrast, the proposed GM-LAMP network can still outperform other beamspace channel estimation schemes.

V. CONCLUSIONS

In this paper, we have proposed a complex-valued knowledge-aided GM-LAMP network to solve the beamspace channel problems in mmWave massive MIMO systems. Due to the sparsity of the beamspace channel, the beamspace channel estimation can be formulated as a sparse channel estimation problem, which can be solved by the LAMP network. By exploiting the Gaussian mixture prior distribution of sparse channel elements, we derive a new shrinkage function, which is different from the original shrinkage function in the existing LAMP network. Based on this new shrinkage function, we propose a complex-valued knowledge-aided GM-LAMP network to estimate the beamspace channel. Simulation results verify that compared with the existing LAMP network and other conventional channel estimation schemes, the proposed GM-LAMP network considering the channel knowledge can achieve better beamspace channel estimation accuracy with a low pilot overhead. We can find that by the leveraging the domain knowledge of the problems to be solved, the general DNN can be redesigned to improve the performance for the specific problems. For the future work, we will follow the idea of the proposed GM-LAMP network to solve the channel estimation problem in terahertz (THz) communications by considering THz channel features.

REFERENCES

- [1] S. Mumtaz, J. Rodriguez, and L. Dai, *MmWave Massive MIMO: A Paradigm for 5G*. New York, NY, USA: Academic, 2016.
- [2] X. Gao, L. Dai, and A. M. Sayeed, "Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 211–217, Apr. 2018.
- [3] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557–1571, Apr. 2016.
- [4] Y. Zeng, R. Zhang, and Z. N. Chen, "Electromagnetic lens-focusing antenna enabled massive MIMO: Performance improvement and cost reduction," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1194–1206, Jun. 2014.
- [5] J. Brady, N. Behdad, and A. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Ant. and Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.
- [6] N. Srinidhi, T. Datta, A. Chockalingam, and B. S. Rajan, "Layered tabu search algorithm for large-MIMO detection and a lower bound on ML performance," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2955–2963, Nov. 2011.
- [7] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 3679–3684.
- [8] P. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2222, Jun. 2015.
- [9] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.
- [10] L. Yang, Y. Zeng, and R. Zhang, "Channel estimation for millimeter-wave MIMO communications with lens antenna arrays," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3239–3251, Apr. 2018.
- [11] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [12] J. Tao, C. Qi, and Y. Huang, "Regularized multipath matching pursuit for sparse channel estimation in millimeter wave massive MIMO system," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 169–172, Feb. 2019.
- [13] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmwave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Jan. 2019.
- [14] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, Feb. 2018.
- [15] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, Sep. 2017.
- [16] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *Proc. Inf. Theory Workshop, Cairo, Egypt, Jan. 2010*, pp. 1–5.
- [17] X. Zou, F. Li, J. Fang, and H. Li, "Computationally efficient sparse bayesian learning via generalized approximate message passing," in *2016 IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB)*, Oct. 2016, pp. 1–4.
- [18] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [19] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast R-CNN for pedestrian detection," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 985–996, Apr. 2018.
- [20] B. Wu, K. Li, F. Ge, Z. Huang, M. Yang, S. M. Siniscalchi, and C. Lee, "An end-to-end deep learning approach to simultaneous speech dereverberation and acoustic modeling for robust speech recognition," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 8, pp. 1289–1300, Dec. 2017.
- [21] F. Liang, C. Shen, and F. Wu, "An iterative BP-CNN architecture for channel decoding," *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 144–159, Feb. 2018.
- [22] M. Kim, N. Kim, W. Lee, and D. Cho, "Deep learning-aided SCMA," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 720–723, Apr. 2018.
- [23] C. Wen, W. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.
- [24] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.
- [25] H. He, C. K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.
- [26] M. Borgerding, P. Schniter, and S. Rangan, "AMP-inspired deep networks for sparse linear inverse problems," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4293–4308, Aug. 2017.
- [27] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [28] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.
- [29] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmwave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sep. 2018.
- [30] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.
- [31] Z. Gao, L. Dai, S. Han, C.-L. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Jun. 2018.