Millimeter Wave Channel Learning Aided by Tensor Completion

Tzu-Hsuan Chou, Nicolo Michelusi, David J. Love, and James V. Krogmeier

Abstract

Index Terms

Millimeter wave, beam-alignment, position-aided, tensor completion, sparse learning.

I. Introduction

Future wireless communication system needs to sustain the condition for higher data rates, lower latency, and better power efficiency [1]. In last few years, millimeter wave (mmWave) communication gains more attentions and be the promising candidate to provide the throughput enhancement for the future wireless network (5G) due to abundant available bandwidth in the spectrum [2]–[5]. The narrow beam property of mmWave is also beneficial for the network quality with the interference reduction. However, the mmWave communication experiences higher path loss than conventional sub-6GHz system, led by air, water, and vapor absorption resulting in the link quality degradation. To conquer the adverse channel condition and to utilize the narrow beam property, it demands the great beamforming gain from the fast and high directional transmission.

The wavelength of mmWave is relatively small that allows the devices packing large number of antennas in small-scale area, especially suitable for Massive MIMO system. The conventional MIMO system using the same number of RF chain as the number of antennas, called digital (baseband) beamforming, may be unsuitable to be applied in massive MIMO system due to the high cost and power consumption of RF chains. To conquer the disadvantage, the analog

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA; emails: {chou59, michelus, djlove, jvk}@purdue.edu.

This research has been funded by NSF under grant CNS-1642982.

beamforming with one RF chain is considered in several works [6], [7]. Analog beamforming ususally uses the phase shifter with constant modulous constraint, such as uniform linear array (ULA) and uniform planar array (UPA). ULA is the phase shifted multiple antennas vector. The system with ULA has been widely studied in **[ref]**. UPA [8], [9] is the array vector packed with the antennas in planar grid, utilizing the 3D beamforming pointing in certain direction with the elevation angle and azimuth angle to provide the beamforming gain, and to eliminate the inter-user interference. **[Hybrid precoding, TBD]**

The traditional channel acquisition in massive MIMO system is unworkable because of the great number of antennas which induce the unacceptibly high channel estimation overhead. It is pessimistic for the massive MIMO because the enormous channel estimation (training) overhead occupies most of the channel coherence time, leading to inadequate amount of time for transmission even it has large gain margin from accurate beamforming. Therefore, the fast and accurate beam-alignment protocol is demanding for enabling the massive MIMO in mmWave communication. The spatial sparsity of mmWave propagation with few dominant clusters (paths) drives a set of beam-alignment scheme on directional transmission. To facilitate the mmWave communication system, the beam-alignment schemes, ranging from beam-sweeping [10], AOA/AOD estimation [11], [12], to data-assisted scheme [13]–[15], are intensely investigated in recent years. Beam-sweeping requires collecting a set of measurements of the possible beam-directions. The exhaustive search is the simplest way belonging to this group, which check the performance of all possible beams. AOA/AOD estimation leverage the spatial sparsity of mmWave channel via compressive sensing technique. **[introduce previous works.]** Regarding the data-assisted scheme, it employs the side information to capture and provide the relation with mmWave channel. The mmWave channel are related to the situation of user, like the user's position, the geometry of environment, the traffic condition, or the accessed base station. **[introduce previous works.]** In [14], it considers the mmWave beam alignment in V2I system with multipath fingerprint which utilizes the long-term channel characteristics associated with the position. In [15], it does the beam recommendation by extracting the information from the prior measurement with the help of position information and environmental situational awareness. It is worth noting that the previous position-aided method develop the beam recommendation algorithm only for the user's positions where the prior measurements are available.

[Problem not yet solved] In all previous beam-alignment method with data-assisted scheme, the database is expected to contain the prior measurements for the entire service area. It requires costly measurement and huge overhead for the data collection, so the assumption of the prior work is inefficient. Once the user is located in the position without prior measurement, the prior works fail to provide accurate recommended beam sets for user.

[Idea of Channel Learning][Matrix completion/ Tensor Completion, TBD]

[Explain my work, TBD]

[Organization:]

II. System Model and Problems Formulation

A. General System Model

We consider a scenario with a base station (BS), serving an area with GPS coordinates $\{(g_x, g_y) : X_0 \leq g_x \leq X_{end}, Y_0 \leq g_y \leq Y_{end}\}$, as in Fig. 1. Considering a uplink MIMO scheme, in which the UE at GPS coordinate $\mathbf{g} = (g_x, g_y)$ sends an $N_t \times 1$ transmit signal vector $\mathbf{x}$, yielding the $N_r \times 1$ received signal vector

$$\mathbf{y}_\mathbf{g} = \mathbf{H}_\mathbf{g}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{H}_\mathbf{g}$ is an $N_r \times N_t$ uplink channel matrix between the BS and the UE at GPS coordinate $\mathbf{g}$, and $\mathbf{n}$ is an N_r-dimensional noise. Assume that the noise to have i.i.d. entries distributed as $\mathcal{CN}(0, \sigma_n^2)$. Here, the channel is related to the position of BS, UE, and clusters in the environment. To the best of our knowledge, this channel model can only be done by real channel measurements or the simulation by ray-tracing software, which simulates the channel measurements from the accurate modeling of the propagation environment. In [16], it studies and states the channel modeling of physical layer based on the propagation environment, called map-based channel model. Our proposed algorithm is applicable to the channel acquired by real measurements, but we generate the channel measurement with the simulator, Quadriga [17], for the evaluation. In the work, the position of the BS and clusters are fixed , so the wireless channel should be strongly related to the UE position.

The system is equipped with analog beamforming vector having one RF chain in both transmitter and receiver. We consider uniform planar arrays (UPAs) at both transmiter and receiver containing N_x and N_y antennas along x and y directions. The inter-antenna element spacing is half-wavelength $\lambda/2$ where λ is the wavelength of the wave propagation. The UPA vector

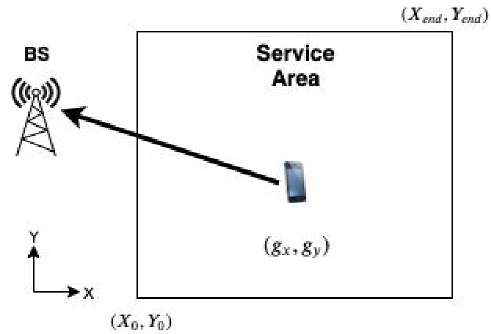

Fig. 1. Communication scenario

$\mathbf{a}(\theta,\phi)$ is the beam steering vector pointing at the direction of the elevation angle and elevation angle (θ,ϕ) expressed as

$$\mathbf{a}(\theta,\phi) = \frac{1}{\sqrt{N_x N_y}}\left[1\ e^{j\Omega_y}\ \cdots e^{j(N_y-1)\Omega_y}\right]^T \otimes \left[1\ e^{j\Omega_x}\ \cdots e^{j(N_x-1)\Omega_x}\right]^T, \tag{2}$$

with $\Omega_y{=}\pi\sin\theta\sin\phi$, $\Omega_x{=}\pi\sin\theta\cos\phi$. The transmit beamforming codebook is the set

$$\mathcal{F}{=}\{\mathbf{f}_{i,j} = \mathbf{a}(\theta_i,\phi_j), i = 1,\cdots,N_\theta, j = 1,\cdots,N_\phi\}$$

of size $|\mathcal{F}|= N_\theta N_\phi$, consisting of the planar array with $N_{tx}\times N_{ty}$ antennas. We define the set of indices corresponding to the beamformers in $\mathcal{F}$ as $\mathcal{I}_f \equiv \{(i_f,j_f) : i_f = 1,\cdots,N_\theta, j_f = 1,\cdots,N_\phi\}$. Similarly, the receive beamforming codebook $\mathcal{W}$ is also constructed as the set of size $|\mathcal{W}|= N_\theta N_\phi$,

$$\mathcal{W}{=}\{\mathbf{w}_{i,j} = \mathbf{a}(\theta_i,\phi_j), i = 1,\cdots,N_\theta, j = 1,\cdots,N_\phi\},$$

consisiting of the planar array with $N_{rx}\times N_{ry}$ antennas. We define the set of indices corresponding to the elements in $\mathcal{W}$ as $\mathcal{I}_w \equiv \{(i_w,j_w) : i_w = 1,\cdots,N_\theta, j_w = 1,\cdots,N_\phi\}$. To construct $\mathcal{F}$ and $\mathcal{W}$, we desire to have the codebook containing the beam directions in the elevation angle $\theta\in[-\pi/2,\pi/2)$ and the azimuth angle $\phi\in[-\pi/2,\pi/2)$, the angle of the codebook θ_i and ϕ_j

are uniformly quantized in $[-\pi/2, \pi/2)$ are

$$\theta_i = -\frac{\pi}{2} + (i-1) \times \frac{\pi}{N_\theta},\ i = 1, \cdots, N_\theta, \tag{3}$$

$$\phi_j = -\frac{\pi}{2} + (j-1) \times \frac{\pi}{N_\phi},\ j = 1, \cdots, N_\phi, \tag{4}$$

where π/N_θ and π/N_ϕ are the resolution of the elevation and azimuth angles. The received signal from the UE at GPS coordinate $\mathbf{g}$ with the beam-pair $(\mathbf{w}_{i,j}, \mathbf{f}_{i,j})$ is expressed by the form

$$\tilde{\mathbf{y}}_{\mathbf{g}}^{\left(\mathbf{b^w}, \mathbf{b^f}\right)} = \sqrt{P_t}\mathbf{w}_{i,j}^\dagger \mathbf{H}_{\mathbf{g}} \mathbf{f}_{i,j} \mathbf{s} + \tilde{\mathbf{n}}, \tag{5}$$

where $\mathbf{b^w} = (i_w, j_w)$ and $\mathbf{b^f} = (i_f, j_f)$ are the indices corresponding to the received and transmit beam codebook; P_t is the transmit power; $\mathbf{s} \in \mathbb{C}^{N \times 1}$ is the known unit-norm transmit signal vector; $\tilde{\mathbf{n}} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$ is the received noise vector due to the unit norm receive beamforming. The received signal strength can be expressed as

$$r_{\mathbf{g}}^{\left(\mathbf{b^w}, \mathbf{b^f}\right)} = \left| \mathbf{s}^\dagger \tilde{\mathbf{y}}_{\mathbf{g}}^{\left(\mathbf{b^w}, \mathbf{b^f}\right)} \right| = \left| \sqrt{P_t}\mathbf{w}_{i,j}^\dagger \mathbf{H}^{\mathbf{g}} \mathbf{f}_{i,j} + \nu \right|. \tag{6}$$

where $\nu = \mathbf{s}^H \tilde{\mathbf{n}}$ is still a zero-mean complex Gaussian noise with variance σ_n^2.

The data collection is the essential step for the machine learning approach to keep the useful information in the database. Here, we describe how we collect data and store the information in the database. We first discretize the service area of the BS with resolution Δ_s and define the position labels $\mathbf{p} = (p_x, p_y)$ as the function of GPS coordinate $\mathbf{g}$,

$$\mathbf{p}(\mathbf{g}) = \left(1 + \left\lfloor \frac{g_x - X_0}{\Delta_s} \right\rceil, 1 + \left\lfloor \frac{g_y - Y_0}{\Delta_s} \right\rceil \right), \tag{7}$$

where $\lfloor x \rceil$ denotes the nearest integer to x; the x-axis label $p_x \in \{1, \cdots, N_x\}$ where $N_x = \left\lceil \frac{X_{end} - X_0}{\Delta_s} \right\rceil$; and the y-axis label $p_y \in \{1, \cdots, N_y\}$ where $N_y = \left\lceil \frac{Y_{end} - Y_0}{\Delta_s} \right\rceil$. In data collection (training) stage, the BS performs the measurement by getting the received signal stength computed with a fixed transmit power P_t. We define the received signal strength with beam pair $(\mathbf{w}_{i,j}, \mathbf{f}_{i,j})$ at position $\mathbf{p}$ as $r^{\left(\mathbf{p}(\mathbf{g}), \mathbf{b^w}, \mathbf{b^f}\right)} = r_{\mathbf{g}}^{\left(\mathbf{b^w}, \mathbf{b^f}\right)}$. During the data collection, the BS might collect multiple measurements for the same beam-pair and position. For beam-pair $(\mathbf{w}_{i,j}, \mathbf{f}_{i,j})$ and position $\mathbf{p}$, we extract the information by computing the average received power,

$$\bar{r}^{(\mathbf{p}, \mathbf{b^w}, \mathbf{b^f})} = \frac{1}{N_{ob}^{(\mathbf{p}, \mathbf{b^w}, \mathbf{b^f})}} \sum_{k=1}^{N_{ob}^{(\mathbf{p}, \mathbf{b^w}, \mathbf{b^f})}} r_k^{(\mathbf{p}, \mathbf{b^w}, \mathbf{b^f})}, \tag{8}$$

where $r_k^{(\mathbf{p},\mathbf{b^w},\mathbf{b^f})}$ is the k-th measured received power and $N_{ob}^{(\mathbf{p},\mathbf{b^w},\mathbf{b^f})}$ is the number of measurements collected so far on that position and beam-pair. Once the BS performs a new measurement for position $\mathbf{p}$ and beam-pair $(\mathbf{w}_{i,j}, \mathbf{f}_{i,j})$, we can online update the average received power as

$$\bar{r}^{(\mathbf{p},\mathbf{b^w},\mathbf{b^f})} \leftarrow \frac{N-1}{N}\bar{r}^{(\mathbf{p},\mathbf{b^w},\mathbf{b^f})} + \frac{1}{N} r_N^{(\mathbf{p},\mathbf{b^w},\mathbf{b^f})},$$

where $N = N_{ob}^{(\mathbf{p},\mathbf{b^w},\mathbf{b^f})}$. Then, the database records the average received power along with the side information, including the UE's position $\mathbf{p} = (p_x, p_y)$, and the indices of the beam-pair codeword $(\mathbf{b^w}, \mathbf{b^f}) = (i_w, j_w, i_f, j_f)$, as in TABLE I.

TABLE I
DATABASE FORM

p_x	p_y	i_w	j_w	i_f	j_f	$\bar{r}^{(\mathbf{p},\mathbf{b^w},\mathbf{b^f})}$
1	1	1	4	1	4	5.2
1	2	4	5	4	5	6.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮

B. What do we do with the cloud's data?

In this section, we discuss how we deal with the statistical channel state information (CSI, average received signal strength) in database, and then provide a general model to formulate the statistical CSI as the function of multiple side information. In equation (8), we observe that the average received signal strength is derived by averaging the measurements with the same position and beam-pair. Each measurement done by the BS is related to the UE position $\mathbf{p} = (p_x, p_y)$ and the beam-pair $(\mathbf{w}_{i,j}, \mathbf{f}_{i,j})$ which UE transmits with. Therefore, we assume that there is a function $G^* : \mathbb{N}^D \rightarrow [0, \infty)$ such that

$$\bar{r}^{(p_x,p_y,i^w,j^w,i^f,j^f)} = G^*(p_x, p_y, i_w, j_w, i_f, j_f), \tag{9}$$

where D is the number of the side information, $D = 6$ in our case. Our goal is to learn the function G from the statistical CSI in the database, which approximates the function G^*.

Learning the function G is similar to charting the radio geometry for the possible positions in the BS service area. The radio geometry reveals the statistical channel information for each position, which facilitates the communication system design, like beam-alignment protocol or channel estimation, etc. The position information of the UE is available via a suite of sensor such as GPS or LIDAR [14], [18]. With the learned function G in our model and the UE's GPS coordinate $\mathbf{g}$, we could extract the radio geometry as $G(\mathbf{p}(\mathbf{g}), :, :, :, :)$, which contains the average received signal strength for all possible beam-pairs. In Section IV, we will provide two application examples utilizing our proposed model to support the communication system design. In [19], it proposes a learning framework in which a multi-antenna system locating the UE according to their received radio geometry. It retrieves the UE position from the received signal, takeing advantage of the relationship between the spatial positions and the radio geometry. Compared with [19], our channel learning model (function learning G) provides a more general form to capture the relationship between the spatial positions and the radio geometry.

If the average received signal strength for all positions, and with all beam measurements available in the database, we learn the function G by mapping each combination of inputs to a corresponding statistical CSI, which successfully approximates the function G^*. However, it is impractical to collect the measurements with all combinations of positions and beam-pairs due to the limited sampling resources. Some positions may have never been observed. Even in the observed positions, there might be only a few number of beams' information recorded in the database. The database would be normally sparse, so it fails the learning of function G. Therefore, we need to predict the received signal strength for the unobserved beams in order to learn the function G. To address this challenge, our objective is to do the prediction for the missing entries in the database with the existing data and properties.

C. Millimeter Wave Example

For the channel matrix $\mathbf{H}$, mmWave channel are expected to have limited scattering [20]. We consider an extended geometric channel model [3] with L multipaths,

$$\mathbf{H} = \sqrt{N_r N_t} \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}_r(\theta_\ell^A, \phi_\ell^A) \mathbf{a}_t^*(\theta_\ell^D, \phi_\ell^D),$$

where $\mathbf{a}_r(\theta_\ell^A, \phi_\ell^A)$ and $\mathbf{a}_t(\theta_\ell^D, \phi_\ell^D)$ are the normalized receive steering vector (see (2)) and the normalized transmit steering vector (see (2)) of the ℓ path; θ_ℓ^A and θ_ℓ^D the elevation angle of arrival and departure; ϕ_ℓ^A and ϕ_ℓ^D the azimuth angle of arrival and departure; α_ℓ is the complex channel gain; N_r (N_t) is the number of receive (transmit) antennas. Due to the spatially sparsity of mmWave channel, there are only few beam pairs leading to significant received power, and the received signal strength of the remaining beam-pairs are negligible. Therefore, in each position, only few beam-pairs have substantial average receive signal strength in the database. In section IV, we will provide two application examples in the scenario of mmWave channel.

III. Tensor Model and Completion

A. Tensor Formulation

A tensor is a multi-dimensional array, i.e. matrix can be looked as a second order tensor. In Section II-B, we formulate the average received signal strength as the function of the UE position and beam-pair. The function G^* in (9) has 6 feature inputs $(p_x, p_y, i^w, j^w, i^f, j^f) \in [1, N_x] \times [1, N_y] \times [1, N_\theta] \times [1, N_\phi] \times [1, N_\theta] \times [1, N_\phi]$, and it is learned by the function G only if the average receive signal strength for all combination of feature inputs are available in the database. A tensor is a suitable data structure to represent the function G, by filling the statistical CSI into the 6-th order tensor $\mathcal{T}_G \in \mathbb{R}^{N_x \times N_y \times N_\theta \times N_\phi \times N_\theta \times N_\phi}$,

$$\mathcal{T}_G(p_x, p_y, i_w, j_w, i_f, j_f) = \bar{r}^{(p_x, p_y, i_w, j_w, i_f, j_f)}. \tag{10}$$

The tensor is complete if there is no missing elements in the database.

Here, we introduce some preliminaries of tensor. An N-th order tensor is defined as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, where the order N of the tensor is the number of dimensions. Given the N-th order tensor $\mathcal{X}$, the $(i_1, i_2, \ldots, i_N)$-th element is $\mathcal{X}(i_1, i_2, \ldots, i_N)$. The inner product of two tensors $\mathcal{X}$ and $\mathcal{Y}$ is defined by

$$\langle \mathcal{X}, \mathcal{Y} \rangle := \sum_{i_1, \cdots, i_N} \mathcal{X}(i_1, \cdots, i_N) \mathcal{Y}(i_1, \cdots, i_N),$$

and the Frobenius norm of a tensor is denoted by

$$\|\mathcal{X}\|_F := \left(\sum_{i_1, \cdots, i_N} |\mathcal{X}(i_1, \cdots, i_N)|^2 \right)^{\frac{1}{2}}.$$

The mode-k unfolding (or the mode-k matricization) of a tensor $\mathcal{X}$ is defined as $\mathcal{X}_{(k)} \in \mathbb{R}^{I_k \times J}$ where $J = \prod_{d=1, d \neq k}^{N} I_d$. The opposite operation of unfolding is defined as $\text{fold}_k(\mathcal{X}_{(k)}) := \mathcal{X}$. In Fig. 2, we demonstrate the visualization of the unfolding operation on a 3-rd order tensor along each mode, and also its corresponding folding operation.

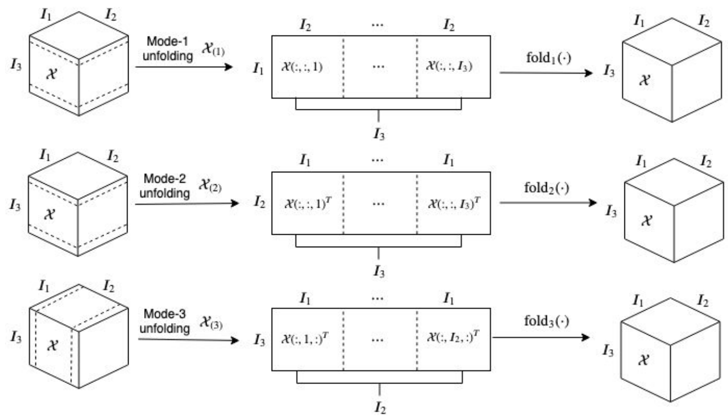

Fig. 2. The visualization of the unfolding and the folding for a 3-rd order tensor $\mathcal{X}$.

B. Tensor Completion

We formulate the problem of learning the function G into the problem of filling the tensor $\mathcal{T}_G$ with the CSI in the database. In section II-B, we address the challenge that it fails to learn the function G from the database because of the deficiency of some combinations of positions and beam-pairs, which means that the data tensor $\mathcal{T}_G$ is an incomplete tensor. Thus, we seek to predict the missing values of the tensor based on the existing information. It is a tensor completion problem.

Matrix/tensor completion is the problem of estimating the missing values using only the available data and the intrinsic properties of the data structure. Here, we review the popular and state-of-art methods for matrix/tensor completions. The widely used completion method is the minimization of the matrix/tensor rank, which captures some useful information of data structure. Considering the matrix $\mathbf{T}$ with Ω as the indices of the known elements, the optimization problem is formulated

$$\min_{\mathbf{X}} \ \mathrm{rank}(\mathbf{X}), \quad \text{s.t. } \mathbf{X}_\Omega = \mathbf{T}_\Omega, \tag{11}$$

where matrix $\mathbf{X}$ is the completed matrix; $\mathbf{X}_\Omega = \mathbf{T}_\Omega$ represents that $\mathbf{X}(i,j) = \mathbf{T}(i,j), \forall (i,j) \in \Omega$ while the remaining elements are missing; $\mathrm{rank}(\mathbf{X})$ represents the rank of matrix $\mathbf{X}$. However, the rank function of $\mathbf{X}$ is a nonconvex function of $\mathbf{X}$, and the rank minimization is an NP-hard problem [21]. To approximate the rank function, the common approach [22] is to minimize the nuclear norm $\|\mathbf{X}\|_*$, the tightest convex envelope for the rank of matrix. The convex optimization problem for the matrix completion is reformulated as

$$\min_{\mathbf{X}} \|\mathbf{X}\|_*, \quad \text{s.t. } \mathbf{X}_\Omega = \mathbf{T}_\Omega, \tag{12}$$

where $\|\mathbf{X}\|_* = \sum_i \sigma_i(\mathbf{X})$ is the nuclear norm of $\mathbf{X}$, and $\sigma_i(\mathbf{X})$ is the i-th largest singular value. The alternating direction method of multipliers (ADMM) [23] is a Lagrangian based approach to efficiently solve the problem (12).

The smoothness property of data is often used in reconstruction problem. If the matrix/tensor is smooth, the differences between the values of neighboring elements in matrix/tensor are comparably small. Considering the incomplete matrix data $\mathbf{T}$ with Ω as the indices of known elements, we formulate the completion problem with smoothness constraint

$$\min_{\mathbf{X}} \ S(\mathbf{X}), \text{ s.t. } \mathbf{X}_\Omega = \mathbf{T}_\Omega, \tag{13}$$

where $\mathbf{X}$ is the completed matrix; $\mathbf{X}_\Omega = \mathbf{T}_\Omega$ represents that $\mathbf{X}(i,j) = \mathbf{T}(i,j), \forall (i,j) \in \Omega$ while the remaining elements are missing; $S(\mathbf{X})$ is the smoothness function of matrix $\mathbf{X}$. In [24], the total variation (TV) is proposed as a smoothness constraint,

$$\|\mathbf{X}\|_{TV} = \sum_{i,j} \sqrt{(\Delta_1 \mathbf{X}(i,j))^2 + (\Delta_2 \mathbf{X}(i,j))^2}, \tag{14}$$

where $\Delta_1 \mathbf{X}(i,j) = \mathbf{X}(i+1,j) - \mathbf{X}(i,j)$ and $\Delta_2 \mathbf{X}(i,j) = \mathbf{X}(i,j+1) - \mathbf{X}(i,j)$. In [25], it exploits the total variation as smoothness constraint $S(\mathbf{X}) = \|\mathbf{X}\|_{TV}$ in reconstruction problem. The completion problem with $\|\mathbf{X}\|_{TV}$ as smoothness constraint is convex, and can be solved by subgradient descend methods. In [26], the matrix completion method with the combination of using the low-rank and smoothness of matrix simultaneously is proposed. It considers a modified linear total variation as the smoothness constraint

$$\|\mathbf{X}\|_{LTV} = \sum_{i,j} (\Delta_1 \mathbf{X}(i,j))^2 + (\Delta_2 \mathbf{X}(i,j))^2 \tag{15}$$

The smooth low-rank matrix completion problem is formulated

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* + \gamma \|\mathbf{X}\|_{LTV}, \text{ s.t. } \mathbf{X}_\Omega = \mathbf{T}_\Omega, \tag{16}$$

where γ is the trade-off factor. This optimization problem is called as the linear total variation approximate regularized nuclear norm minimization problem (LTVNN), and able to be solved by ADMM method.

The low-rank tensor completion is the natural extension problem of low-rank matrix completion. Given a N-th order incomplete tensor $\mathcal{T}$ with Ψ as the indices of known elements, the problem is formulated as

$$\begin{aligned} &\min_{\mathcal{X}} \ \text{rank}(\mathcal{X}) \\ &s.t. \ \mathcal{X}_\Psi = \mathcal{T}_\Psi \end{aligned} \tag{17}$$

The big issue is about the definition of the nuclear norm of the tensor. Since computing the rank of a tensor is an NP-hard problem [27], there is no expression for the convex envelope of the tensor rank. In [28], it proposed the following definition for the nuclear norm of the tensor,

$$\|\mathcal{X}\|_* = \sum_{k=1}^{N} \alpha_k \|\mathcal{X}_{(k)}\|_* \tag{18}$$

where the constant coefficients $\alpha_k > 0$ and $\sum_{k=1}^{N} \alpha_k = 1$. The proposed nuclear norm of the tensor is convex since it is a linear combination of the nuclear norms of all unfolding matrices along each modes. In [28], the high accuracy low-rank tensor completion (HaLRTC) method was proposed on the idea of minimizing the tensor nuclear norm defined in (18), and the problem is formulated

$$\begin{aligned} &\min_{\mathcal{X}, \mathcal{Y}_1, \ldots, \mathcal{Y}_N} \ \sum_{k=1}^{N} \alpha_k \|(\mathcal{Y}_k)_{(k)}\|_* \\ &\text{s.t. } \mathcal{X}_\Psi = \mathcal{T}_\Psi, \ \mathcal{X} = \mathcal{Y}_k. \end{aligned} \tag{19}$$

It minimizes the nuclear norm of the unfolding matrices of the tensor $\mathcal{X}$ along each mode. This problem can be formulated in to the Lagrange function and solved by ADMM method.

C. Tensor Completion Algorithm with Noise-free Data

In previous section, we review the state-of-art method for matrix/tensor completion supported by the intrinsic properties of data. However, the tensor properties are not always consistent through all modes since it depends on the feature selection of the inputs. Considering the data tensor $\mathcal{T}_G$ in (10), $\mathcal{T}_G$ is a 6-th order tensor containing the elements of average received signal strength corresponding to 6 different features, where the first 2 features are the position information and the last 4 features are the beam information. For a give beam-pair index $(\hat{i}_w, \hat{j}_w, \hat{i}_f, \hat{i}_f)$, the subtensor $\mathcal{T}_G(:, :, \hat{i}_w, \hat{j}_w, \hat{i}_f, \hat{i}_f) \in \mathbb{R}^{N_x \times N_y}$ is a second-order tensor (matrix) with smoothness property. The smoothness originates from the fact that the received signal strength varies smoothly between neighboring positions on a given beam-pair. It is related to the position resolution Δ_s defined in (7). For the fixed position $(\hat{p}_x, \hat{p}_y)$, the subtensor $\mathcal{T}_G(\hat{p}_x, \hat{p}_y, :, \cdots, :) \in \mathbb{R}^{N_\theta \times N_\phi \times N_\theta \times N_\phi}$ is the tensor with low-rank property if the syetem is operated in mmWave channel. It comes from the fact that the received signal strength in the same position concentrates in few beam clusters due to the limited scattering of mmWave channel.

Definition 1. The generalize linear total variation $GLTV(\mathcal{X})$ of the tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, which quantifies the smoothness of $\mathcal{X}$, is defined as

$$GLTV(\mathcal{X}) = \sum_{i_1,\ldots,i_N} \sum_{k=1}^{N} (\Delta_k \mathcal{X}(i_1, \cdots, i_N))^2,$$

where $\Delta_k \mathcal{X}(i_1, \cdots, i_N) = \mathcal{X}(i_1, \cdots, i_k + 1, \cdots, i_N) - \mathcal{X}(i_1, \cdots, i_k, \cdots, i_N)$. It is the generalized form of the linear total variation (see (15)).

On top of that, I am providing the algorithm for completing the tensor with different data properties in distinct feature dimensions. Considering the tensor $\mathcal{T} \in \mathbb{R}^{I_1^s \times \cdots \times I_{n_1}^s \times I_1^\ell \times \cdots \times I_{n_2}^\ell}$, we separate the dimensions of $\mathcal{T}$ into first n_1 dimensions with smooth property and last n_2 dimensions with low-rank property. It means that the n_1-th order subtensor with the same indices of last n_2 dimensions, $\mathcal{T}(:, \cdots, :, i_1^\ell, \cdots, i_{n_2}^\ell) \in \mathbb{R}^{I_1^s \times \cdots \times I_{n_1}^s}$, is a smooth tensor; and the n_2-th order subtensor with the same indices of first n_1 dimensions, $\mathcal{T}(i_1^s, \cdots, i_{n_1}^s, :, \cdots, :) \in \mathbb{R}^{I_1^\ell \times \cdots \times I_{n_2}^\ell}$, is low-rank. We assume that the tensor $\mathcal{T}$ is an incomplete tensor with observed indices set Ψ, which means the element of $\mathcal{T}$ in observed set Ψ are given while the remaining elements are missing. To do the estimation for the missing elements in tensor $\mathcal{T}$, we formulate the optimization problem as

$$\begin{aligned} \arg\min_{\mathcal{X}} & \left(\sum_{i_1^s,\cdots,i_{n_1}^s} \|\mathcal{X}(i_1^s, \cdots, i_{n_1}^s, :, \cdots, :)\|_* \right) \\ & + \gamma \left(\sum_{i_1^\ell,\cdots,i_{n_2}^\ell} S\left(\mathcal{X}(:, \cdots, :, i_1^\ell, \cdots, i_{n_2}^\ell)\right) \right) \\ \text{s.t. } & \mathcal{X}_\Psi = \mathcal{T}_\Psi, \end{aligned} \tag{20}$$

where γ is the trade-off parameter. The first term of the objective function is the sum of nuclear norm of all possible low-rank subtensors $\mathcal{X}(i_1^s, \cdots, i_{n_1}^s, :, \cdots, :) \in \mathbb{R}^{I_1^\ell \times \cdots \times I_{n_2}^\ell}$, $\forall (i_1^s, \cdots, i_{n_1}^s) \in [1, I_1^s] \times \cdots \times [1, I_{n_1}^s]$. The second term of the objective function is the sum of smoothness function of all possible subtensors $\mathcal{X}(:, \cdots, :, i_1^\ell, \cdots, i_{n_2}^\ell) \in \mathbb{R}^{I_1^s \times \cdots \times I_{n_2}^s}$, $\forall (i_1^\ell, \cdots, i_{n_1}^\ell) \in [1, I_1^\ell] \times$

$\cdots \times [1, I_{n_1}^{\ell}]$. Here, we consider $S(\mathcal{X}) = GLTV(\mathcal{X})$ defined in *Definition* 1.

We use the alternating direction method of multipliers (ADMM) [29] to efficiently solve (20). About the nuclear norm of the tensor $\|\mathcal{X}(i_1^s, \cdots, i_{n_1}^s, :, \cdots, :)\|_*$, we consider the proposed tensor nuclear norm in (18). With ADMM and HaLRTC in (19), we reformulate the optimization problem as

$$
\begin{aligned}
&\left\{\mathcal{X}, \{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}, \mathcal{B}\right\} = \\
&\arg\ \min\ \left(\sum_{i_1^s,\cdots,i_{n_1}^s} \sum_{k=1}^{n_2} \alpha_k \left\| \left\{ \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} \right\}_{(k)} \right\|_* \right) \\
&+ \gamma \left(\sum_{i_1^\ell,\cdots,i_{n_2}^\ell} S\left(\mathcal{B}(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell)\right) \right) \\
&+ \sum_{i_1^s,\cdots,i_{n_1}^s} \sum_{k=1}^{n_2} \frac{\lambda}{2} \left\| \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} - \mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) \right\|_F^2 \\
&+ \frac{\mu}{2} \|\mathcal{B} - \mathcal{X}\|_F^2 \\
&\text{s.t. } \mathcal{X}_\Psi = \mathcal{T}_\Psi, \mathcal{X} = \mathcal{B}, \\
&\quad \mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) = \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}, \\
&\quad k = 1,\cdots,n_2,\ \forall (i_1^s,\cdots,i_{n_1}^s),
\end{aligned} \tag{21}
$$

where λ and μ are small fixed positive parameters; the constant coefficients $\alpha_k > 0$ and $\sum_{k=1}^{n_2} \alpha_k = 1$. We introduce two different kind of Lagrangian multipliers: $\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}$ associated with the constraint $\mathcal{X}(i_1^s, \cdots, i_{n_1}^s, :, \cdots, :) = \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}$ for $k = 1, \cdots, n_2$ and $(i_1^s, \cdots, i_{n_1}^s) \in [1, I_1^s] \times \cdots \times [1, I_{n_1}^s]$; and $\mathcal{Z}$ associated with the constraint $\mathcal{X} = \mathcal{B}$. The augmented Lagrange function of (21) is expressed as

$$
\begin{aligned}
& L\left(\mathcal{X},\left\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\right\},\mathcal{B},\left\{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\right\},\mathcal{Z}\right)\\
=& \sum_{i_1^s,\cdots,i_{n_1}^s}\sum_{k=1}^{n_2}\left\{\alpha_k\left\|\left\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\right\}_{(k)}\right\|_*\right.\\
& +\frac{\lambda}{2}\left\|\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}-\mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:)\right\|_F^2\\
& \left.+\left\langle\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)},\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}-\mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:)\right\rangle\right\}\\
& +\left\{\gamma\sum_{i_1^\ell,\cdots,i_{n_2}^\ell}S\left(\mathcal{B}(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell)\right)\right.\\
& \qquad\left.+\frac{\mu}{2}\|\mathcal{B}-\mathcal{X}\|_F^2+\langle\mathcal{Z},\mathcal{B}-\mathcal{X}\rangle\right\}.
\end{aligned}
\tag{22}
$$

The ADMM is implemented by minimizing the augmented Lagrange function L over $\mathcal{X}$, $\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}$, $\mathcal{B}$ and then update the Lagrange multiplier $\{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}$, $\mathcal{Z}$ as in (23). To optimize $\mathcal{X}$, we minimize the Lagrange function L w.r.t. $\mathcal{X}$ with fixed $\left\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\right\}$, $\mathcal{B}$, $\left\{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\right\}$, and $\mathcal{Z}$, yielding the

---

$$
\begin{aligned}
\mathcal{X}_{t+1} &= \arg\min_{\mathcal{X}} L\left(\mathcal{X},\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_t,\mathcal{B}_t,\{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_t,\mathcal{Z}_t\right),\quad \text{s.t. } \mathcal{X}_\Psi=\mathcal{T}_\Psi;\\
\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_{t+1} &= \arg\min_{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}} L\left(\mathcal{X}_{t+1},\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)},\mathcal{B}_t,\{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_t,\mathcal{Z}_t\right),\ \forall k=1,\cdots,n_2;\\
\mathcal{B}_{t+1} &= \arg\min_{\mathcal{B}} L\left(\mathcal{X}_{t+1},\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_{t+1},\mathcal{B},\{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_t,\mathcal{Z}_t\right);\\
\{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_{t+1} &= \{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_t+\beta_1\left(\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_{t+1}-\mathcal{X}_{t+1}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:)\right),\ \forall k=1,\cdots,n_2;\\
\mathcal{Z}_{t+1} &= \mathcal{Z}_t+\beta_2\left(\mathcal{B}_{t+1}-\mathcal{X}_{t+1}\right);
\end{aligned}
\tag{23}
$$

---

following optimization problem,

$$
\begin{aligned}
\min_{\mathcal{X}} & \sum_{i_1^s,\cdots,i_{n_1}^s} \sum_{k=1}^{n_2} \left\{ \frac{\lambda}{2} \left\| \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} - \mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) \right\|_F^2 \right. \\
& \left. + \left\langle \mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}, \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} - \mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) \right\rangle \right\} \\
& + \frac{\mu}{2} \|\mathcal{B} - \mathcal{X}\|_F^2 + \langle \mathcal{Z}, \mathcal{B} - \mathcal{X} \rangle \\
\text{s.t. } & \mathcal{X}_\Psi = \mathcal{T}_\Psi.
\end{aligned} \tag{24}
$$

By computing the derivative of (24) with respect to the element of $\mathcal{X}$ for the unknown entries $\mathcal{X}(\psi), \forall \psi \notin \Psi$ and setting them equal to zero, the solution is given by

$$
\begin{aligned}
& \hat{\mathcal{X}}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) = \\
& \frac{1}{\mu + n_2\lambda} \left\{ \sum_{k=1}^{n_2} \left(\lambda \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} + \mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)} \right) \right. \\
& \qquad \left. + (\mu\mathcal{B} + \mathcal{Z})(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) \right\},
\end{aligned} \tag{25}
$$

$\forall (i_1^s,\cdots,i_{n_1}^s)$. We can update the tensor $\mathcal{X}$ by

$$
\mathcal{X}(\psi) = \begin{cases} \hat{\mathcal{X}}(\psi), & \psi \notin \Psi, \\ \mathcal{T}(\psi), & \psi \in \Psi. \end{cases} \tag{26}
$$

For the minimization of L over $\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}$ with the remaining parameters fixed, the problem can be formulated as

$$
\begin{aligned}
& \arg \min_{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}} \\
& \sum_{i_1^s,\cdots,i_{n_1}^s} \sum_{k=1}^{n_2} \left\{ \frac{\alpha_k}{\lambda} \left\| \left(\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} \right)_{(k)} \right\|_* + \frac{1}{2} \left\| \left(\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} \right)_{(k)} \right. \right. \\
& \left. \left. - \left(\mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) - \frac{1}{\lambda} \mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)} \right)_{(k)} \right\|_F^2 \right\}.
\end{aligned}
$$

For each $(i_1^s,\cdots,i_{n_1}^s)$ and k, the problem is strictly convex [22] and the solution is given by singular value thresholding. Thus, the update of $\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}$ can be written as

$$\begin{aligned}&\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} = \\ &\text{fold}_k\Big(\mathcal{D}_{\frac{\alpha_k}{\lambda}}\Big(\Big(\mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) - \frac{1}{\lambda}\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\Big)_{(k)}\Big)\Big),\end{aligned} \tag{27}$$

where $\mathcal{D}_\tau$ is the soft-thresholding operator. For a matrix $\mathbf{A}$ with singular value decomposition $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, where $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \ldots \sigma_r)$, this is defined as $\mathcal{D}_\tau(\mathbf{A}) = \mathbf{U}\mathcal{D}_\tau(\boldsymbol{\Sigma})\mathbf{V}^H,\ \mathcal{D}_\tau(\boldsymbol{\Sigma}) = \text{diag}(\{\max\{\sigma_i - \tau, 0\}\})$.

To optimize $\mathcal{B}$, we minimize the Lagrange function L w.r.t $\mathcal{B}$ with all the remaining variables fixed. The optimization problem is formulated as

$$\begin{aligned}&\arg\min_{\mathcal{B}} \sum_{i_1^\ell,\cdots,i_{n_2}^\ell} \Big\{\gamma S\left(\mathcal{B}(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell)\right) \\ &+ \frac{\mu}{2}\left\|(\mathcal{B}-\mathcal{X})(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell)\right\|_F^2 \\ &+ \Big\langle \mathcal{Z}(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell), (\mathcal{B}-\mathcal{X})(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell)\Big\rangle\Big\}.\end{aligned} \tag{28}$$

We solve this optimization problem for each $(i_1^\ell,\cdots,i_{n_2}^\ell)$ separately. Since the smoothness function $S(\cdot)$ (GLTV) is a quadratic function of $\mathcal{B}(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell)$, the objective function is also a quadratic function of $\mathcal{B}(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell)$. For a given $(i_1^\ell,\cdots,i_{n_2}^\ell)$, by computing the derivative of (28) with respect to the elements in subtensor $\mathcal{B}_s = \mathcal{B}(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell) \in \mathbb{R}^{I_1^s\times\cdots\times I_{n_1}^s}$ and setting it to zero, yielding the equations, $\forall(i_1^s,\cdots,i_{n_1}^s) \in [1, I_1^s]\times\cdots\times[1, I_{n_1}^s]$,

$$\Big\langle \mathcal{U}_{(i_1^s,\cdots,i_{n_1}^s)}, \mathcal{B}_s\Big\rangle - \frac{(\mu\mathcal{X}-\mathcal{Z})(i_1^s,\cdots,i_{n_1}^s,i_1^\ell,\cdots,i_{n_2}^\ell)}{2\gamma} = 0, \tag{29}$$

where $\mathcal{U}_{(i_1^s,\cdots,i_{n_1}^s)} \in \mathbb{R}^{I_1^s\times\cdots\times I_{n_1}^s}$ is defined as

$$\begin{aligned}&\left(\mathcal{U}_{(i_1^s,\cdots,i_{n_1}^s)}\right)(\alpha_1,\cdots,\alpha_{n_1}) = \\ &\begin{cases} \sum_{k=1}^{n_1}\delta(i_k^s>1)+\delta(i_k^s<I_k^s) & \\ +\frac{\mu}{2\gamma} & \text{, if } \sum_{k=1}^{n_1}|\alpha_k - i_k^s| = 0 \\ -1 & \text{, if } \sum_{k=1}^{n_1}|\alpha_k - i_k^s| = 1 \\ 0, & \text{, otherwise.}\end{cases}\end{aligned}$$

According to (29), there are $\prod_{k=1}^{n_1} I_k^s$ unknowns (the elements in $\mathcal{B}_s$) and $(\prod_{k=1}^{n_1} I_k^s)$ linear equations, so the tensor $\mathcal{B}_s$ can be derived by solving (29), then $\mathcal{B}$ is updated. Finally, we update the Lagrangian multiplier $\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}$ and $\mathcal{Z}$ as in (23). The algorithm updates $\mathcal{X}$, $\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}$, $\mathcal{B}$, $\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}$, and $\mathcal{Z}$ iteratively until the stop criterion is satisfied, shown in **Algorithm 1**. With the convergence of ADMM [29], the iteration approaches the primal feasibility $\Big\| \mathcal{X}_t(i_1^s, \cdots, i_{n_1}^s, :, \cdots, :) - \{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s}$ 0 and $\|\mathcal{X}_t - \mathcal{B}_t\|_F \to 0$, thus we set the stop criterion as $\epsilon < \epsilon_0$, where

$$\begin{aligned}\epsilon = \sum_{i_1^s,\cdots,i_{n_1}^s} \sum_{k=1}^{n_2} \left\| \mathcal{X}(i_1^s, \cdots, i_{n_1}^s, :, \cdots, :) - \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} \right\|_F \\ + \|\mathcal{X} - \mathcal{B}\|_F. \qquad (30)\end{aligned}$$

and ϵ_0 is a threshold. With the primal feasibilities, we can induce that

$$\begin{aligned}&\left\| \{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_{t+1} - \{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_t \right\|_F \\ &= \beta_1 \left\| \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} - \mathcal{X}(i_1^s, \cdots, i_{n_1}^s, :, \cdots, :) \right\|_F \to 0,\end{aligned}$$

and

$$\|\mathcal{Z}_{t+1} - \mathcal{Z}_t\|_F = \beta_2 \|\mathcal{B} - \mathcal{X}\|_F \to 0,$$

which guarantee the convergence of Lagrange multipliers $\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}$ and $\mathcal{Z}$.

Algorithm 1 Tensor Completion with Noise-free Data

Input: incomplete data tensor $\mathcal{T}$, observed set Ψ

Output: $\hat{\mathcal{T}}$

Initialization $\mathcal{X} = \mathcal{T}$, $\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} = \mathcal{T}(i_1^s, \cdots, i_{n_1}^s, :, \cdots, :)$, $\mathcal{B} = \mathcal{T}$, $\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)} = 0$, $\mathcal{Z} = 0$, and $\epsilon = \infty$.
while $\epsilon > \epsilon_0$ **do**
    Update $\mathcal{X}_{t+1}$ by (26)
    Update $\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_{t+1},\ \forall k,\ \forall (i_1^s, \cdots, i_{n_1}^s)$, by (27)
    Update $\mathcal{B}_{t+1}$ by solving (29)
    Update $\{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\}_{t+1},\ \forall k,\ \forall (i_1^s, \cdots, i_{n_1}^s)$, by (23)
    Update $\mathcal{Z}_{t+1}$ by (23)
    Calculate ϵ by (30); $t := t + 1$
end while
$\hat{\mathcal{T}} = \mathcal{X}$

D. Tensor Completion Algorithm with Noisy Data [TBD]

As we know, the noise exists in the practical wireless communication. The recorded data in the database is naturally contaminated by the noise, which comes from the environment or the hardware of communication system. Therefore, we should consider this noise constraint of the data in tensor completion.

$$\begin{aligned} \arg\min_{\mathcal{X}} & \left(\sum_{i_1^s,\cdots,i_{n_1}^s} \|\mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:)\|_* \right) \\ & + \gamma \left(\sum_{i_1^\ell,\cdots,i_{n_2}^\ell} S\left(\mathcal{X}(:,\cdots,:,i_1^\ell,\cdots,i_{n_2}^\ell)\right) \right) \\ \text{s.t. } & f_1(\mathcal{X}) \leq 0. \end{aligned} \tag{31}$$

where the constraint function is defined as

$$f_1(\mathcal{X}) = \sum_{\psi\in\Psi} \left| \{\mathcal{X}_\Psi - \mathcal{T}_\Psi\}(\psi) \right|^2 - d,$$

and $d = g(|\Psi|, \sigma_n^2)$ is the upperbound of the sum of the noise terms. It is an optimization problem with an inequality constraint. We could also apply the ADMM to solve this problem as below. The ADMM is implemented by minimizing the augmented Lagrange function L over $\mathcal{X}, \left\{\mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)}\right\}, \mathcal{B}$ and then update the Lagrange multiplier $\left\{\mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}\right\}, \mathcal{Z}$. Compared with the optimization problem of noise-free case, the update of $\mathcal{X}$ is the only different part of the solving procedure, and all the remaining subproblems are the same. The subproblem for updating $\mathcal{X}$ is described as

$$\begin{aligned} \min_{\mathcal{X}} & \sum_{i_1^s,\cdots,i_{n_1}^s} \sum_{k=1}^{n_2} \left\{ \frac{\lambda}{2} \left\| \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} - \mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) \right\|_F^2 \right. \\ & \left. + \left\langle \mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}, \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} - \mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) \right\rangle \right\} \\ & + \frac{\mu}{2} \|\mathcal{B} - \mathcal{X}\|_F^2 + \langle \mathcal{Z}, \mathcal{B} - \mathcal{X} \rangle \\ \text{s.t. } & f_1(\mathcal{X}) \leq 0. \end{aligned}$$

To solve this problem, we utilize the barrier method by adding $\left(-\frac{1}{t}\right) log(-f_1(\mathcal{X}))$ as the penalty term to the objective function,

$$
\begin{aligned}
\min_{\mathcal{X}} \sum_{i_1^s,\cdots,i_{n_1}^s} \sum_{k=1}^{n_2} \Bigg\{ & \frac{\lambda}{2} \left\| \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} - \mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) \right\|_F^2 \\
& + \left\langle \mathcal{Y}_k^{(i_1^s,\cdots,i_{n_1}^s)}, \mathcal{A}_k^{(i_1^s,\cdots,i_{n_1}^s)} - \mathcal{X}(i_1^s,\cdots,i_{n_1}^s,:,\cdots,:) \right\rangle \Bigg\} \\
& + \frac{\mu}{2} \left\| \mathcal{B} - \mathcal{X} \right\|_F^2 + \langle \mathcal{Z}, \mathcal{B} - \mathcal{X} \rangle + \left(-\frac{1}{t}\right) log(-f_1(\mathcal{X})).
\end{aligned}
$$

E. Complexity Analysis ***[TBD]***

IV. EXAMPLE APPLICATIONS **[TBD]**

In this section, we provide the examples which apply our proposed channel learning method in communication system. Here, we consider two examples of problems in mmWave channel and illustrate how our method can help. For both examples, we assume that the completed tensor $\hat{\mathcal{T}}$, derived by our proposed tensor completion method, is given. In first example, we consider the analog beamforming system with UPA codebook at both the transmitter and the receiver. We would like to design a algorithm providing few and accurate recommendations of beam-pairs with the help of the statistical CSI predicted by our proposed method. In second example, we consider the hybrid precoding communication system. We aim to provide the sensing matrix from the statistical CSI provided by our learning method. We consider the assumption that the comple

A. Beam recommendation for analog precoding

The problem of the example is to provide a set of candidate beam-pairs at a give UE position. The overhead of the conventional beam-sweeping approach scales up with the product of the receive codebook size $|\mathcal{W}|$ and the transmit codebook size $|\mathcal{F}|$, which is unacceptibly large. We aim to propose a learning algorithm which provides a small beam-pair set $\mathcal{S} = \{(\mathbf{w}, \mathbf{f}) : \mathbf{w} \in \mathcal{W},\ \mathbf{f} \in \mathcal{W}\}$, which is likely to contain the best beam-pair (the one with the largest received signal strength).

In Fig. 3, we introduce a flow diagram of the position-aided beam-pair recommendation protocol. In step 1, the UE initiates the uplink mmWave transmission request with its GPS coordinate $\mathbf{g} = (g_x, g_y)$ to the BS using sub-6GHz control channels. In step 2, the BS forwards the GPS information to the cloud, which processes the learning algorithm to provide the recommended subset of beam-pairs. Then, the BS conveys the recommended beam-pair set to the UE. In step 3, it trains all the candidate beam-pairs in the recommended subset by sending a sequence of $|\mathcal{S}|$ signals from the UE to the BS with the corresponding beamforing codewords in the recommeded set. In step 4, the BS selects the best beam-pair $(\hat{\mathbf{b}}_{\mathbf{w}}, \hat{\mathbf{b}}_{\mathbf{f}})$ according to the received signal strength, and then feedback the codeword index of the best transmit beam index $\hat{\mathbf{b}}_{\mathbf{f}}$ to the UE. Finally, the UE and the BS use the selected beam-pair $(\hat{\mathbf{b}}_{\mathbf{w}}, \hat{\mathbf{b}}_{\mathbf{f}})$ to do the subsequent mmWave transmission. Due to the reciprocity of the mmWave channel, the selected beam-pair can also be used for downlink transmission but in different direction.

[How can our framework help] Define the beam-pair set

$$\mathcal{K} = \{(\mathbf{b}_{\mathbf{w}}, \mathbf{b}_{\mathbf{f}}) : \mathbf{b}_{\mathbf{w}} \in \mathcal{I}_w, \;\; \mathbf{b}_{\mathbf{f}} \in \mathcal{I}_f\}$$

[Proposed Algorithm]

Algorithm 2 Beam Subset Selection

Input: completed tensor $\hat{\mathcal{T}}$, beam number N_{tr}, receive/transmit codebook beam-pair indices $\mathcal{K}$, UE position $\mathbf{p} = (p_x, p_y)$

Output: recommended beam subset $\mathcal{S}_{N_{tr}}$

Initialization $\mathcal{S}_0 \leftarrow \emptyset$

for $n = 1 : N_{tr}$ **do**

$\quad (\mathbf{b}_{\mathbf{w}}{}^*, \mathbf{b}_{\mathbf{f}}{}^*) = \arg\max_{(\mathbf{b}_{\mathbf{w}}, \mathbf{b}_{\mathbf{f}}) \in \mathcal{K} \setminus \mathcal{S}_{n-1}} \hat{\mathcal{T}}(\mathbf{p}, \mathbf{b}_{\mathbf{w}}, \mathbf{b}_{\mathbf{f}})$

$\quad \mathcal{S}_n \leftarrow \mathcal{S}_{n-1} \cup (\mathbf{b}_{\mathbf{w}}{}^*, \mathbf{b}_{\mathbf{f}}{}^*)$

end for

B. Sensing matrix recommendation for hybrid precoding

[Explain the example]

[Hybrid precoder system model]

[How can our framework help]

[Proposed Algorithm]

V. Online Tensor Completion **[TBD]**

[With warm-start, or smaller size.]

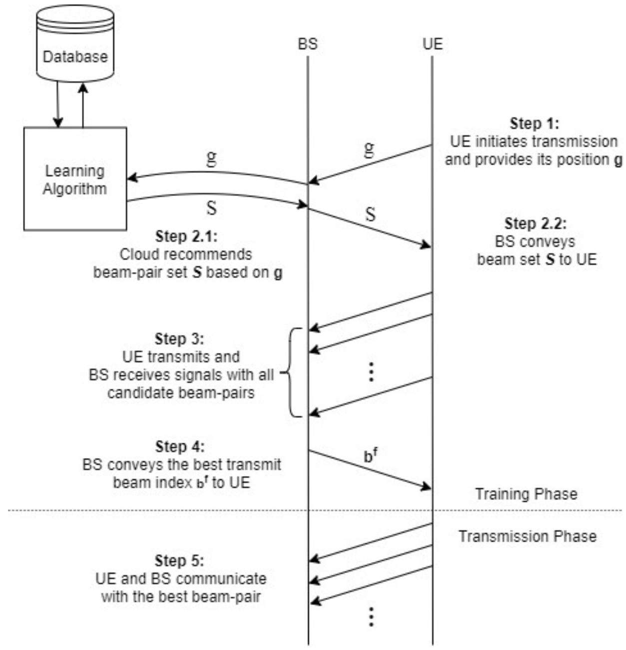

Fig. 3. Position-aided beam recommendation

VI. Numerical Results [TBD]

A. Simulation setup

We consider uplink MIMO, with an UPA having $N_x = 16$ and $N_y = 16$ antennas along the x and y directions (as in (2)) at both the BS and the UE. The scenario *mmMAGIC_UMi_NLOS* is selected, with carrier frequency $f_c = 58.68$ GHz. The receive (transmit) UPA codebook size is $|\mathcal{W}|= 256$ ($|\mathcal{F}|= 256$) with $(N_\theta, N_\phi) = (16, 16)$. The network layout is in Fig. 4, containing one BS at $(0, 0, 10)$ serving the UE in the area $\mathcal{A} = \{(g_x, g_y) : 10 \leq g_x \leq 60, -25 \leq g_y \leq 25\}$ with height as 1.5 m. Considering $51 \times 51 = 1261$ reference GPS coordinates uniformly located in the service area $\mathcal{A}$, we collect the MIMO channel for each reference GPS coordinate in $\mathcal{A}$ as the ground truth data. The position labels of $\mathcal{A}$ are derived as in (7) with the resolution $\Delta_s = 5$m, where the length are $N_x = 11$ and $N_y = 11$. The data tensor can be expressed by $\mathcal{T} \in \mathbb{R}^{N_x \times N_y \times N_\theta \times N_\phi \times N_\theta \times N_\phi}$, where $(N_x, N_y, N_\theta, N_\phi) = (11, 11, 16, 16)$. The observed position ratio $K_{op} = C_{op}/(L_x L_y)$ is varied, where C_{op} denotes the number of observed positions. Regarding the observed set of data tensor, we make the two following assumptions for the experiment setting. **Assusmption 1:** The observed positions are randomly chosen. For the observed position $\mathbf{p}'$, the measurements of the reference GPS coordinates corresponding to position $\mathbf{p}'$, $\{\mathbf{g} = (g_x, g_y) : \mathbf{p}(\mathbf{g}) = \mathbf{p}',\ \mathbf{g} \in \mathcal{A}\}$, are observed. **Assusmption 2:** For each observed GPS coordinate

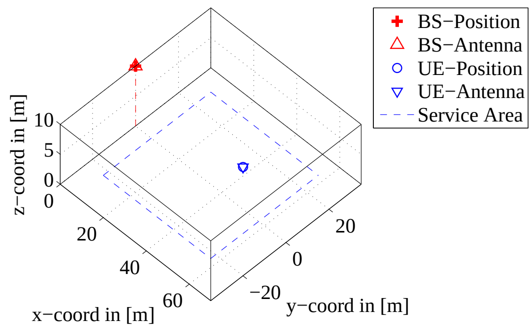

Fig. 4. The network layout of the scenario

g, only the measurements of the top 10% beams (ranked by received signal strength) are stored in the database. With these two assumptions, the data tensor is incomplete in both positions' and beams' dimensions.

B. Evaluation of the average signal strength prediction

[MSE of predicted received signal strength, for Noise-free case and Noisy case]

C. Beam recommendation for analog beamforming

[The Beam recommendation example]

D. Sensing matrix recommendation for hybrid precoder

[The sensing matrix recommendation example]

E. Time-varying channel

[Example for time-varying channel, the propagation environment changes due to mobility.]

F. Online learning

[Example for online learning.]

VII. CONCLUSIONS

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5g be?” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] T. S. Rappaport *et al.*, “Millimeter wave mobile communications for 5g cellular: It will work!” *IEEE Access*, vol. 1, pp. 335–349, 2013.

[3] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter wave mimo systems,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, April 2016.

[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive mimo for next generation wireless systems,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[5] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5g,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.

[6] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, “Millimeter wave beamforming for wireless backhaul and access in small cell networks,” *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, October 2013.

[7] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, “Mimo for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both?” *IEEE Communications Magazine*, vol. 52, no. 12, pp. 110–121, December 2014.

[8] Y. Nam, B. L. Ng, K. Sayana, Y. Li, J. Zhang, Y. Kim, and J. Lee, “Full-dimension mimo (fd-mimo) for next generation cellular technology,” *IEEE Communications Magazine*, vol. 51, no. 6, pp. 172–179, June 2013.

[9] S. Noh, M. D. Zoltowski, and D. J. Love, “Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[10] M. Hussain and N. Michelusi, “Energy-efficient interactive beam alignment for millimeter-wave networks,” *IEEE Trans. on Wireless Communications*, vol. 18, no. 2, pp. 838–851, Feb 2019.

[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. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct 2014.

[12] Z. Marzi, D. Ramasamy, and U. Madhow, “Compressive channel estimation and tracking for large arrays in mm-wave picocells,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, April 2016.

[13] A. Ali and R. W. Heath, “Compressed beam-selection in millimeterwave systems with out-of-band partial support information,” in *Proc. of the IEEE International Conf. on Acoustics, Speech and Signal Processing,*, March 2017, pp. 3499–3503.

[14] V. Va, J. Choi, T. Shimizu, G. Bansal, and R. W. Heath, “Inverse multipath fingerprinting for millimeter wave v2i beam alignment,” *IEEE Trans. Veh. Technol*, vol. 67, no. 5, pp. 4042–4058, May 2018.

[15] Y. Wang, M. Narasimha, and R. W. Heath, “Mmwave beam prediction with situational awareness: A machine learning approach,” in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.

[16] 3GPP, “Study on channel model for frequencies from 0.5 to 100 GHz ,” 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.901, 10 2019, version 16.0.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3173

[17] S. Jaeckel *et al.*, "Quadriga: A 3-d multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 6, pp. 3242–3256, June 2014.

[18] E. Ward and J. Folkesson, "Vehicle localization with low cost radar sensors," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, June 2016, pp. 864–870.

[19] C. Studer, S. Medjkouh, E. Gonultas, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.

[20] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, April 2013.

[21] N. Gillis and F. Glineur, "Low-rank matrix approximation with weights or missing data is np-hard," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1149–1165, Nov. 2011. [Online]. Available: http://dx.doi.org/10.1137/110820361

[22] J. feng Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optimization*, 2010.

[23] J. Yang and X. Yuan, "Linearized augmented lagrangian and alternating direction methods for nuclear norm minimization," *Mathematics of Computation*, vol. 82, no. 281, pp. 301–329, 2013.

[24] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992. [Online]. Available: https://doi.org/10.1016/0167-2789(92)90242-F

[25] Q. Dai and W. Sha, "The physics of compressive sensing and the gradient-based recovery algorithms," *CoRR*, vol. abs/0906.1487, 2009. [Online]. Available: http://arxiv.org/abs/0906.1487

[26] X. Han, J. Wu, L. Wang, Y. Chen, L. Senhadji, and H. Shu, "Linear total variation approximate regularized nuclear norm optimization for matrix completion," *Abstract and Applied Analysis*, p. 8, 2014.

[27] C. J. Hillar and L.-H. Lim, "Most tensor problems are np-hard," *J. ACM*, vol. 60, no. 6, pp. 45:1–45:39, Nov. 2013. [Online]. Available: http://doi.acm.org/10.1145/2512329

[28] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan 2013.

[29] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.