

Learning on a Grassmann Manifold: CSI Quantization for Massive MIMO Systems

Keerthana Bhogi, Chiranjib Saha, and Harpreet S. Dhillon

Abstract—This paper focuses on the design of beamforming codebooks that maximize the average normalized beamforming gain for any underlying channel distribution. While the existing techniques use statistical channel models, we utilize a model-free data-driven approach with foundations in machine learning to generate beamforming codebooks that adapt to the surrounding propagation conditions. The key technical contribution lies in reducing the codebook design problem to an unsupervised clustering problem on a Grassmann manifold where the cluster centroids form the finite-sized beamforming codebook for the channel state information (CSI), which can be efficiently solved using K -means clustering. This approach is extended to develop a remarkably efficient procedure for designing product codebooks for full-dimension (FD) multiple-input multiple-output (MIMO) systems with uniform planar array (UPA) antennas. Simulation results demonstrate the capability of the proposed design criterion in learning the codebooks, reducing the codebook size and producing noticeably higher beamforming gains compared to the existing state-of-the-art CSI quantization techniques.

Index Terms—Massive MIMO, FD-MIMO, FDD, beamforming, codebook, machine learning, Grassmann manifold, K -means clustering.

I. INTRODUCTION

Transmit beamforming with receive combining is one of the simplest approaches to achieve full diversity in a MIMO system. It just requires CSI at the transmitter in the form of the transmit beamforming vector. The frequency division duplex (FDD) large-scale MIMO systems cannot utilize channel reciprocity to acquire CSI at the transmitter using uplink transmission. This necessitates channel estimation using downlink pilots and the subsequent feedback of the channel estimates (in this case the beamforming vector) to the transmitter over a dedicated feedback channel with limited capacity. This results in a significant overhead when the number of antennas is large. One way to overcome this problem is to construct a set of beamforming vectors constituting a *codebook*, which is known to both the transmitter and the receiver. The problem then reduces to determining the *best* beamforming vector at the receiver and conveying its *index* to the transmitter over the feedback channel [1]. A key step in this procedure is to construct the codebook, which is a classical problem in MIMO communications [2]. A common feature of all classical works in this direction is the assumption of a statistical channel model (such as Rayleigh) for which the *optimal* codebooks are constructed to optimize system performance.

With the recent interest in data-driven approaches to wireless system design, it is quite natural to wonder whether machine learning has any role to play in this classical problem. Since the fundamental difficulty in this problem is the dimensionality of the channel, the natural tendency is to think in terms of obtaining a low dimensional representation of the channel using deep learning techniques, such as autoencoders, and use it for codebook construction. An autoencoder operates on the hypothesis that the data possesses a representation on a lower dimensional manifold (referred to as *feature space*), *albeit* unknown, and tries to learn the embedded manifold by training over the dataset. In contrast, for MIMO beamforming, the underlying manifold is known to be a Grassmann manifold. This removes the requirement of “learning” the manifold from the dataset which often times can be extremely complicated. Once the manifold is known, we can leverage the “shallow” learning techniques like the clustering algorithms on the manifold to find the optimal codebook for beamforming.

Prior Art. As is the case with any communication theory problem, almost all existing works on limited feedback assume *some* analytical model for the channel to enable tractable analyses, e.g., i.i.d Rayleigh fading [3], spatial correlation [4], temporal correlation [5] or both [6]. Specifically, the problem of quantized maximum ratio transmission (MRT) beamforming can be interpreted as a Grassmannian line packing problem for both uncorrelated [7] and spatially correlated [8] Rayleigh fading channels and has been extensively studied. The idea of connecting Grassmann manifolds to wireless communications is not new and has been used in other aspects of MIMO systems, such as non-coherent communication [9] and limited feedback unitary precoding [10]. Coming to the context of the limited feedback FDD-MIMO, the codebook based on Grassmannian line packing is strictly dependent on the assumption of Rayleigh fading and hence cannot be extended to more realistic scenarios. On the other hand, the discrete Fourier transform (DFT) based beamforming exploits the second order statistics of the channel (such as the direction of departure of the dominant path) and offers a simple yet robust solution to the codebook construction. Owing to the direct connection with the spatial parameters of the channel, the DFT codebook can be extended to the Kronecker product (KP) codebook for 3D beamforming in FD-MIMO scenarios. A major drawback of DFT codebook is that it scans all possible directions even though many of them may not be used and thus the available feedback bits are not used efficiently. Finally, since we will be proposing a clustering based solution, it is useful to note that clustering has already found applications in many related problems, such as MIMO detection [11], automatic

K. Bhogi, C. Saha, and H. S. Dhillon are with Wireless@VT, Department of ECE, Virginia Tech, Blacksburg, VA, USA. Email: {kbhogi, csaha, hdhillon}@vt.edu. The support of the US National Science Foundation (Grants ECCS-1731711 and CNS-1923807) is gratefully acknowledged.

modulation recognition [12], and radio resource allocation in a heterogeneous network [13].

Contributions. The key technical contribution of this paper lies in the novel formulation of transmit beamforming codebook design for any arbitrary channel distribution as the *Grassmannian K -means clustering* problem. First, we develop the algorithm for K -means clustering on the Grassmann manifold that finds the K centroids of the clusters. Leveraging the fact that optimal MRT beamforming vectors lie on a Grassmann manifold, we then develop the design criterion for optimal beamforming codebooks. We then formally establish the connection between the Grassmannian K -means algorithm and the codebook design problem and show that the optimal codebook is nothing but the set of centroids given by the K -means algorithm. This approach is further extended to develop product codebooks for FD-MIMO systems employing UPA antennas. In particular, we show that under the rank-1 approximation of the channel, the optimal codebook can be decomposed as the Cartesian product of two Grassmannian codebooks of smaller dimensions. We discuss the optimality and performance of the codebooks using both the proposed techniques in terms of average normalized beamforming gain.

Notation. We use boldface small case (upper case) letters, e.g. $\mathbf{a}(\mathbf{A})$, to designate column vectors (matrices) with entries in \mathbb{C} . We use $\mathbb{C}^{M \times N}$ to represent $M \times N$ dimensional complex space, $\mathcal{U}(M, N)$ to represent the set of all $M \times N$ orthonormal matrices. Further, $a^*(\mathbf{a}^*)$ denotes complex conjugate of $a \in \mathbb{C}$ ($\mathbf{a} \in \mathbb{C}^{M \times 1}$), \mathbf{A}^T denotes transpose, \mathbf{A}^H denotes hermitian, $\text{svd}(\mathbf{A})$ denotes the singular value decomposition (SVD), $\text{vec}(\mathbf{A})$ denotes the vectorization of a matrix \mathbf{A} . Also, $\mathbb{E}_{\mathbf{a}}[\cdot]$ represents the expectation over the distribution of \mathbf{a} , $|\cdot|$ denotes the absolute value, $\|\cdot\|_2$ denotes the matrix two-norm and $j = \sqrt{-1}$.

II. SYSTEM OVERVIEW

We consider a narrowband point-to-point $M_t \times M_r$ MIMO communication scenario, where the transmitter and receiver are equipped with M_t and M_r antennas, respectively. In this paper, we focus on the transmit beamforming operation, where the transmitter sends one data stream over a flat fading channel. The discrete-time baseband input-output relation for this system can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{f}s + \mathbf{n}, \quad (1)$$

where $\mathbf{y} \in \mathbb{C}^{M_r \times 1}$ is the received baseband signal, $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ is the block fading MIMO channel, $s \in \mathbb{C}$ is the transmitted symbol, $\mathbf{n} \in \mathbb{C}^{M_r \times 1}$ is the additive noise at the receiver, and $\mathbf{f} \in \mathbb{C}^{M_t \times 1}$ is the beamforming vector. The symbol energy is given by $\mathbb{E}[|s|^2] = \mathcal{E}_t$ and the total transmitted energy is $\mathbb{E}[\|\mathbf{f}s\|_2^2] = \mathcal{E}_t \|\mathbf{f}\|_2^2$. The additive noise \mathbf{n} is Gaussian, i.e., entries in \mathbf{n} are i.i.d according to $\mathcal{CN}(0, N_o)$. It is assumed that perfect channel knowledge is always available at the receiver. With the combining vector $\mathbf{z} \in \mathbb{C}^{M_r \times 1}$, the estimated transmitted symbol is obtained as $\hat{s} = \mathbf{z}^H \mathbf{y}$. The receive SNR is

$$\gamma_r = \frac{\mathcal{E}_t |\mathbf{z}^H \mathbf{H} \mathbf{f}|^2}{|\mathbf{z}^H \mathbf{n} \mathbf{n}^H \mathbf{z}|} = \gamma_t \frac{|\mathbf{z}^H \mathbf{H} \mathbf{f}|^2}{\|\mathbf{z}\|_2^2 \|\mathbf{f}\|_2^2},$$

where $\gamma_t = \mathcal{E}_t \|\mathbf{f}\|^2 / N_o$ is the transmit SNR. Without loss of generality, it is assumed that $\|\mathbf{z}\|_2^2 = 1$. Under this assumption

$$\gamma_r = \gamma_t \frac{|\mathbf{z}^H \mathbf{H} \mathbf{f}|^2}{\|\mathbf{f}\|^2} = \gamma_t \Gamma(\mathbf{f}, \mathbf{z}), \quad (2)$$

where $\Gamma(\mathbf{f}, \mathbf{z})$ is the effective channel gain or the beamforming gain. The MIMO beamforming problem is to choose \mathbf{f} and \mathbf{z} such that $\Gamma(\mathbf{f}, \mathbf{z})$ is maximized, which would in turn maximize the SNR and consequently minimize the average probability of error and maximize the capacity [14]. A receiver that employs maximum ratio combining (MRC) chooses \mathbf{z} such that $\Gamma(\mathbf{f}, \mathbf{z})$ for a given \mathbf{f} is maximized [7]. Under the assumption that receiver always uses MRC, \mathbf{z} is given by

$$\mathbf{z} = \mathbf{H}\mathbf{f} / \|\mathbf{H}\mathbf{f}\|_2,$$

and $\Gamma(\mathbf{f}, \mathbf{z})$ can be simplified as

$$\Gamma := \Gamma(\mathbf{f}) = \|\mathbf{H}\mathbf{f}\|_2^2. \quad (3)$$

Therefore the MIMO beamforming problem is to find the optimal beamforming vector \mathbf{f} that maximizes $\Gamma(\mathbf{f})$ and can be formally posed as

$$\mathbf{f} = \arg \max_{\mathbf{x} \in \mathbb{C}^{M_t \times 1}} \Gamma(\mathbf{x}). \quad (4)$$

To constrain transmit power, we assume that $\|\mathbf{f}\|_2^2 = 1$ without loss of generality. We consider MRT, which selects \mathbf{f} to maximize $\Gamma(\mathbf{f}, \mathbf{z})$ for a given \mathbf{z} [15]. Under the assumptions of MRT, receive MRC, and no other design constraints on \mathbf{f} , for a given N_o and \mathcal{E}_t , the optimal beamforming vector \mathbf{f} that maximizes Γ is

$$\begin{aligned} \mathbf{f} &= \arg \max_{\mathbf{x} \in \mathbb{C}^{M_t \times 1}} \|\mathbf{H}\mathbf{x}\|_2^2 \text{ subjected to } \|\mathbf{x}\|_2^2 = 1 \\ &= \arg \max_{\mathbf{x} \in \mathcal{U}(M_t, 1)} \|\mathbf{H}\mathbf{x}\|_2^2. \end{aligned} \quad (5)$$

Note that $\arg \max$ of any function returns only one out of its possibly many global maximizers and thus the output may not necessarily be unique. For an MRT system, \mathbf{f} is the orthonormal eigenvector associated with the maximum eigenvalue of $\mathbf{H}^H \mathbf{H}$ [16]. Let $\lambda_1 \geq \dots \geq \lambda_{M_t}$ be the eigenvalues of $\mathbf{H}^H \mathbf{H}$ and $\mathbf{v}_1, \dots, \mathbf{v}_{M_t}$ be the corresponding eigenvectors. One possible solution of (5) is $\mathbf{f} = \mathbf{v}_1$ and the corresponding beamforming gain is

$$\Gamma(\mathbf{v}_1) = \|\mathbf{H}\mathbf{v}_1\|_2^2 = \lambda_1. \quad (6)$$

For a given \mathbf{H} , let the solution space of (5) be denoted as $\mathcal{S}_{\mathbf{H}} \subset \mathcal{U}(M_t, 1)$. Then $\mathbf{v}_1 \in \mathcal{S}_{\mathbf{H}}$ and for every $\mathbf{f} \in \mathcal{S}_{\mathbf{H}}$, $\Gamma(\mathbf{f}) = \lambda_1$.

Quite obviously, MRT beamforming requires channel state information at the transmitter (CSIT). In particular, in an FDD system, the receiver estimates the channel \mathbf{H} and sends \mathbf{v}_1 back to the transmitter over a feedback channel. Thus the feedback overhead increases as M_t increases. Since the feedback channel is typically assumed to be a low-rate reliable channel, it is not always possible to transmit \mathbf{v}_1 over this channel without any data compression [1]. One way to model this feedback bottleneck is to assume the feedback channel to be a zero-delay and error-free with its capacity being limited to B bits per channel use. Thus, it is necessary to introduce

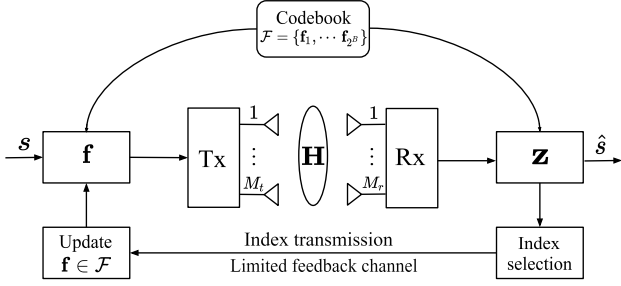


Fig. 1: Block diagram of an FDD-MIMO system with limited feedback channel of capacity B bits per channel use.

some method of quantization for \mathbf{v}_1 . The most well-known approach for the quantization is to construct a dictionary of beams [1], also known as the *beam codebook*. In particular, the transmitter and receiver agree upon a finite set of possible beamforming vectors, say $\mathcal{F} = \{\mathbf{f}_1, \dots, \mathbf{f}_{2^B}\}$ of cardinality 2^B . The receiver chooses the appropriate vector $\mathbf{f} \in \mathcal{F}$ that maximizes Γ and feeds the index of the codeword back to the transmitter. The system-level diagram of a limited feedback FDD-MIMO system, as discussed so far, is provided in Fig. 1. Therefore for a given codebook \mathcal{F} , the optimal beamforming vector as stated in (4) is

$$\mathbf{f} = \arg \max_{\mathbf{f}_i \in \mathcal{F}} \|\mathbf{H}\mathbf{f}_i\|_2^2. \quad (7)$$

It is important to note that the original problem of finding the optimal MRT solution for (5) is a constrained optimization problem on the Euclidean space $\mathbb{C}^{M_t \times 1}$ and does not have unique solution on $\mathbb{C}^{M_t \times 1}$. This problem can be reformulated as a manifold optimization problem as follows. As argued in [7], it can be shown that the optimal MRT beamformers for every $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ lie on a special kind of Riemann manifold embedded in $\mathbb{C}^{M_t \times 1}$, known as the Grassmann manifold. This will be discussed in Section III. As it will be evident later, this manifold structure of the search domain of \mathbf{f} in (5) is the key enabler for our data-driven codebook design.

III. CLUSTERING ON A GRASSMANN MANIFOLD

In this section, we provide a brief introduction to the Grassmann manifold, which is instrumental for the design of the proposed beamforming codebook design. The reader is referred to the foundational texts in differential geometry, such as [17] and tutorials, such as [18], for a comprehensive and rigorous treatment of this manifold. A Grassmann manifold refers to a set of subspaces embedded in a higher-dimensional space (such as the surface of a sphere in \mathbb{R}^3). More formally, the complex Grassmann manifold $\mathcal{G}(M_t, M)$ is defined as

$$\mathcal{G}(M_t, M) := \{\text{span}(\mathbf{Y}) : \mathbf{Y} \in \mathbb{C}^{M_t \times M}, \mathbf{Y}^H \mathbf{Y} = \mathbf{I}_M\}. \quad (8)$$

Any element \mathcal{Y} in $\mathcal{G}(M_t, M)$ is typically represented by an orthonormal matrix \mathbf{Y} whose columns span \mathcal{Y} . It is to be noted that there exists no unique representation of a subspace \mathcal{Y} . This can be explained as follows. Let \mathbf{Y} be the orthonormal basis that spans \mathcal{Y} , then \mathcal{Y} can also be spanned by some other orthonormal matrix $\mathbf{Y}' = \mathbf{Y}\mathbf{R}$ for some $\mathbf{R} \in \mathcal{U}(M, M)$. Thus \mathbf{Y} and \mathbf{Y}' span the same subspace, which is represented by an equivalence relation $\mathbf{Y}' \equiv \mathbf{Y}$. Each of these M -dimensional

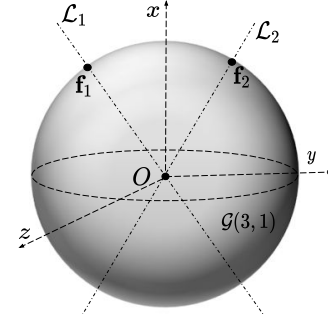


Fig. 2: Example of a Grassmann manifold $\mathcal{G}(3, 1)$ embedded in \mathbb{R}^3 : $\mathcal{L}_1, \mathcal{L}_2$ are two arbitrary lines passing through the origin in \mathbb{R}^3 . They are represented by unit vectors $\mathbf{f}_1, \mathbf{f}_2$, at the intersections of the lines $\mathcal{L}_1, \mathcal{L}_2$ and the surface of the unit sphere in \mathbb{R}^3 .

linear subspaces can be regarded as a single point on the Grassmann manifold, which is represented by its orthonormal basis. Since a linear subspace can be specified by an arbitrary basis, each point on $\mathcal{G}(M_t, M)$ is an equivalence class of orthonormal matrices. Specifically, \mathbf{Y} and $\mathbf{Y}\mathbf{R}$ correspond to the same point on $\mathcal{G}(M_t, M)$.

Now consider the case of $M = 1$, i.e. $\mathcal{G}(M_t, 1)$, which is the set of all one-dimensional subspaces in $\mathbb{C}^{M_t \times 1}$. In other words, one can visualize $\mathcal{G}(M_t, 1)$ as the collection of all lines passing through the origin of the space $\mathbb{C}^{M_t \times 1}$. A line \mathcal{L} passing through origin in $\mathbb{C}^{M_t \times 1}$ is represented in $\mathcal{G}(M_t, 1)$ by a unit vector \mathbf{f} that spans the line. It can also be generated by any other unit vector \mathbf{f}' if $\mathbf{f}' \equiv \mathbf{f}$. Also note, \mathbf{f} and \mathbf{f}' correspond to the same point on $\mathcal{G}(M_t, 1)$. Since a complex Grassmann of arbitrary dimensions is difficult to visualize, for illustration purpose, we present the real Grassmann manifold $\mathcal{G}(3, 1)$ in \mathbb{R}^3 in Fig. 2.

One way of defining the “distance” function on $\mathcal{G}(M_t, 1)$ is by letting the distance between the lines generated by two unit vectors $\mathbf{f}_1, \mathbf{f}_2$ be the sine of the angle between the lines [19]. This distance is expressed as

$$d(\mathbf{f}_1, \mathbf{f}_2) = \sin(\theta_{1,2}) = \sqrt{1 - |\mathbf{f}_1^H \mathbf{f}_2|^2}. \quad (9)$$

The connection between $\mathcal{G}(M_t, 1)$ and optimal MRT beamforming is established next.

Lemma 1. *For a given channel realization $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ and $\mathbf{f} \in \mathcal{S}_{\mathbf{H}}$, every \mathbf{f}' such that $\mathbf{f}' \equiv \mathbf{f}$ is also an optimal MRT beamformer.*

Proof: Given $\mathbf{f} \in \mathcal{S}_{\mathbf{H}}$ and $\mathbf{f}' \equiv \mathbf{f}$ i.e. $\mathbf{f}' = \mathbf{f}e^{j\theta}$ for some $\theta \in [0, 2\pi)$, then $\|\mathbf{H}\mathbf{f}\|_2^2 = |\mathbf{f}^H \mathbf{H}^H \mathbf{H} \mathbf{f}| = |e^{-j\theta} \mathbf{f}^H \mathbf{H}^H \mathbf{H} \mathbf{f} e^{j\theta}| = |\mathbf{f}'^H \mathbf{H}^H \mathbf{H} \mathbf{f}'| = \|\mathbf{H}\mathbf{f}'\|_2^2$. Therefore, \mathbf{f}' is also an element in $\mathcal{S}_{\mathbf{H}}$. ■

Remark 1. *According to Lemma 1, every $\mathbf{f}, \mathbf{f}' \in \mathcal{S}_{\mathbf{H}}$ such that $\mathbf{f}' \equiv \mathbf{f}$ correspond to the same point on $\mathcal{G}(M_t, 1)$. Therefore, any probability distribution on channel $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ will impose a probability distribution on \mathbf{v}_1 in $\mathcal{G}(M_t, 1)$. Thus we need a quantizer defined on $\mathcal{G}(M_t, 1)$ to encode the optimal MRT beamforming vectors and generate a B -bit beamforming codebook $\mathcal{F} \subset \mathcal{G}(M_t, 1)$ for limited feedback beamforming.*

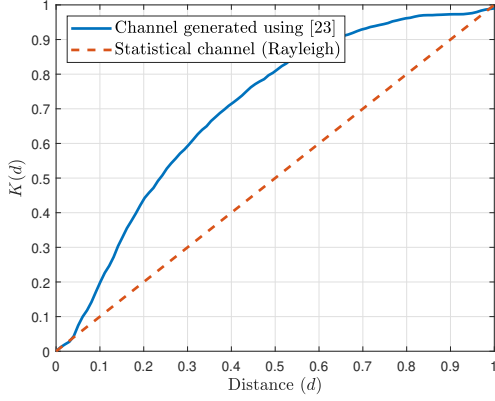


Fig. 3: Comparison of Ripley's K function $K(d)$ of different channel models with $M_t = 2$ and $M_r = 1$.

The optimal MRT beamformers are points on $\mathcal{G}(M_t, 1)$, hence can be described as a *point process* whose characteristics depend on the underlying channel distribution.

Remark 2. For a Rayleigh fading channel where the entries of \mathbf{H} are i.i.d according to $\mathcal{CN}(0, 1)$, it has been shown in [7] that \mathbf{v}_1 is uniformly distributed on $\mathcal{G}(M_t, 1)$. Thus, the construction of \mathcal{F} is equivalent to finding the best packing of 2^B lines in $\mathcal{G}(M_t, 1)$ [19]. See [7] for more details.

For an arbitrary distribution of $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$, the distribution of \mathbf{v}_1 will no longer be uniformly distributed on $\mathcal{G}(M_t, 1)$. As an illustrative example, using the distance function defined in (9), we compare the Ripley's K function $K(d)$ [20] of the optimal MRT beamforming vectors on $\mathcal{G}(M_t, 1)$ for a realistic scenario (see Section VI for more details on the experimental setup) with Rayleigh fading channels for the same underlying system model. Ripley's K function is a spatial descriptive statistic that measures the deviation of a point process from spatial homogeneity. From Fig. 3, we see that the distribution of the optimal MRT beamforming vectors for the realistic channel is significantly different from the uniform distribution on $\mathcal{G}(M_t, 1)$ which is equivalent to *complete randomness* on the manifold. We can also infer that optimal MRT beamformers of the channel for the considered scenario exhibit clustering tendency on $\mathcal{G}(M_t, 1)$. Therefore, a reasonable codebook construction scheme is to simply deploy an unsupervised clustering method (such as K -means clustering) on $\mathcal{G}(M_t, 1)$ that can identify optimal $K = 2^B$ cluster centroids and form the codebook \mathcal{F} . In what follows, we will establish that K -means clustering on $\mathcal{G}(M_t, 1)$ actually yields an optimal codebook.

A. Grassmannian K -means Clustering

K -means clustering on a given metric space is a method of vector quantization to partition a set of n data points into K non-overlapping clusters in which each data point belongs to the cluster with the nearest cluster centroid. The centroids are the quantized representations of the data points that belong to the respective clusters. A quantizer on the given metric space maps the data points to one of the K centroids. The K centroids are chosen such that the average distortion due to the quantization according to a pre-defined distortion measure

is minimized. Before we formally introduce the main steps of the clustering algorithm on $\mathcal{G}(M_t, 1)$, we first define the notion of a distortion measure and a quantizer as follows.

Definition 1 (Distortion measure). The distortion caused by representing $\mathbf{f} \in \mathcal{G}(M_t, 1)$ with $\mathbf{f}' \in \mathcal{G}(M_t, 1)$ is defined as the distortion measure d_o which is given by $d_o(\mathbf{f}, \mathbf{f}') = d^2(\mathbf{f}, \mathbf{f}')$.

Definition 2 (Grassmann quantizer). Let $\mathcal{F} \subset \mathcal{G}(M_t, 1)$ be a B -bit codebook such that $\mathcal{F} = \{\mathbf{f}_1, \dots, \mathbf{f}_{2^B}\}$, then a Grassmann quantizer $Q_{\mathcal{F}}$ is defined as a function mapping elements of $\mathcal{G}(M_t, 1)$ to elements of \mathcal{F} i.e. $Q_{\mathcal{F}} : \mathcal{G}(M_t, 1) \mapsto \mathcal{F}$.

A performance measure of a Grassmann quantizer is the average distortion $D(Q_{\mathcal{F}})$, where

$$D(Q_{\mathcal{F}}) := \mathbb{E}_{\mathbf{x}} [d_o(\mathbf{x}, Q_{\mathcal{F}}(\mathbf{x}))] = \mathbb{E}_{\mathbf{x}} [d^2(\mathbf{x}, Q_{\mathcal{F}}(\mathbf{x}))]. \quad (10)$$

In most practical settings, we may have access to a set of n data points $\mathcal{X} = \{\mathbf{x}\}$ in lieu of the probability distribution $p(\mathbf{x})$. Then the expectation w.r.t \mathbf{x} in (10) means averaging over the set \mathcal{X} . Therefore the objective of K -means clustering with $K = 2^B$ is to find the set of K centroids, i.e. \mathcal{F}^K , that minimizes $D(Q_{\mathcal{F}})$ and can be expressed as

$$\mathcal{F}^K = \arg \min_{\substack{\mathcal{F} \subset \mathcal{G}(M_t, 1) \\ |\mathcal{F}| = 2^B}} D(Q_{\mathcal{F}}) = \arg \min_{\substack{\mathcal{F} \subset \mathcal{G}(M_t, 1) \\ |\mathcal{F}| = 2^B}} \mathbb{E}_{\mathbf{x}} [d^2(\mathbf{x}, Q_{\mathcal{F}}(\mathbf{x}))], \quad (11)$$

and the associated quantizer is

$$Q_{\mathcal{F}^K}(\mathbf{x}) = \arg \min_{\mathbf{f}_i \in \mathcal{F}} d_o(\mathbf{x}, \mathbf{f}_i) = \arg \min_{\mathbf{f} \in \mathcal{F}} d^2(\mathbf{x}, \mathbf{f}). \quad (12)$$

However, finding the optimal solution for K -means clustering is an NP-hard problem. Therefore, we use the Linde-Buzo-Gray algorithm [21] (outlined in Alg. 1) which is a heuristic algorithm that iterates between updating the cluster centroids and mapping a data point to the corresponding centroid that guarantees convergence to a local optimum. In Alg. 1, the only non-trivial step is the centroid calculation for a set of points. In contrast to the squared distortion measure in the Euclidean domain, the centroid of n elements in a general manifold with respect to an arbitrary distortion measure does not necessarily exist in a closed form. However, the centroid computation on $\mathcal{G}(M_t, 1)$ is feasible because of the following Lemma.

Lemma 2 (Centroid computation). For a set of points $\mathcal{V}_k = \{\mathbf{x}_i\}_{i=1}^{N_k}$, $\mathbf{x}_i \in \mathcal{G}(M_t, 1)$, that form the k -th Voronoi partition, the centroid \mathbf{f}_k is

$$\mathbf{f}_k = \arg \min_{\mathbf{f} \in \mathcal{G}(M_t, 1)} \sum_{i=1}^{N_k} d^2(\mathbf{x}_i, \mathbf{f}) = \text{eig} \left(\sum_{i=1}^{N_k} \mathbf{x}_i \mathbf{x}_i^H \right), \quad (14)$$

where $\text{eig}(\mathbf{Y})$ is the dominant eigenvector of the matrix \mathbf{Y} .

IV. GRASSMANNIAN CODEBOOK DESIGN

In this section, we formally establish the connection between Grassmannian K -means clustering and the optimal codebook construction. The transmitter and receiver use a B -bit codebook $\mathcal{F} \subset \mathcal{G}(M_t, 1)$ to map the channel matrix \mathbf{H} to a codeword in $\mathbf{f} \in \mathcal{F}$ according to (7). In order to define

Algorithm 1 Grassmannian K -means Algorithm for (11)

```

1: procedure CODEBOOK( $\mathcal{X}, K$ )
2:   Initialize random  $\mathcal{F} = \{\mathbf{f}_1, \dots, \mathbf{f}_K\}$  on  $\mathcal{G}(M_t, 1)$ 
3:    $\mathcal{V}_k \leftarrow \{\mathbf{x} : d(\mathbf{x}, \mathbf{f}_k) \leq d(\mathbf{x}, \mathbf{f}_j), \forall \mathbf{x} \in \mathcal{X}, k \neq j\} \forall k \in \{1, \dots, K\}$ 
4:    $Q_{\mathcal{F}}(\mathbf{x}) \leftarrow \arg \min_{\mathbf{f} \in \mathcal{F}} d^2(\mathbf{x}, \mathbf{f}) \forall \mathbf{x} \in \mathcal{X}$ 
5:   while ! stopping criteria do
6:     Solve for  $\mathcal{F}$ :

      $\mathbf{f}_k \leftarrow \arg \min_{\mathbf{f}} \sum \mathbf{d}^2(\mathbf{x}, \mathbf{f}) \forall \mathbf{x} \in \mathcal{V}_k, \forall k \in \{1, \dots, K\}$ 

      $Q_{\mathcal{F}}(\mathbf{x}) \leftarrow \arg \min_{\mathbf{f} \in \mathcal{F}} d^2(\mathbf{x}, \mathbf{f}) \forall \mathbf{x} \in \mathcal{X}$ 

      $\mathcal{V}_k \leftarrow \{\mathbf{x} : d(\mathbf{x}, \mathbf{f}_k) \leq d(\mathbf{x}, \mathbf{f}_j), \forall \mathbf{x} \in \mathcal{X}, k \neq j\} \forall k \in \{1, \dots, K\}$ 

   return  $\mathcal{F}$ 

```

the *optimality* of the codebook, we first introduce the average normalized beamforming gain for \mathcal{F} as

$$\begin{aligned} \Gamma_{av} &:= \mathbb{E}_{\mathbf{H}} \left[\frac{\Gamma(\mathbf{f})}{\Gamma(\mathbf{v}_1)} \right] = \mathbb{E}_{\mathbf{H}} \left[\frac{\|\mathbf{H}\mathbf{f}\|_2^2}{\lambda_1} \right] \\ &= \mathbb{E}_{\mathbf{H}} \left[\sum_{i=1}^{M_t} \frac{\lambda_i |\mathbf{v}_i^H \mathbf{f}|^2}{\lambda_1} \right] \geq \mathbb{E}_{\mathbf{v}_1} \left[|\mathbf{v}_1^H \mathbf{f}|^2 \right]. \end{aligned}$$

To measure the average distortion introduced by the quantization using \mathcal{F} , we use the loss in Γ_{av} as given below:

$$\begin{aligned} L(\mathcal{F}) &:= \mathbb{E}_{\mathbf{H}} [1 - \Gamma_{av}] \\ &\leq \mathbb{E}_{\mathbf{v}_1} [1 - |\mathbf{v}_1^H \mathbf{f}|^2] := L_{ub}(\mathcal{F}), \end{aligned} \quad (15)$$

where $L_{ub}(\mathcal{F})$ is an upper bound of $L(\mathcal{F})$ and the sufficient condition for equality in (15) is $\text{rank}(\mathbf{H}) = 1$. The optimal codebook intends to minimize $L(\mathcal{F})$ by minimizing its upper bound $L_{ub}(\mathcal{F})$ as given in (15). Since the current limited feedback approach quantizes \mathbf{v}_1 as \mathbf{f} , it is reasonable to minimize $L_{ub}(\mathcal{F})$ which depends only on \mathbf{v}_1 and \mathbf{f} . This yields the following codebook design criterion.

Definition 3 (Codebook design criterion). *Over all of the B -bit codebooks $\mathcal{F} \subset \mathcal{G}(M_t, 1)$, the Grassmannian codebook \mathcal{F}^* is the one that minimizes $L_{ub}(\mathcal{F})$. Therefore*

$$\mathcal{F}^* := \arg \min_{\substack{\mathcal{F} \subset \mathcal{G}(M_t, 1) \\ |\mathcal{F}| = 2^B}} L_{ub}(\mathcal{F}). \quad (16)$$

Building on this discussion, we now state the method to find the optimal codebook in $\mathcal{G}(M_t, 1)$ as follows.

Theorem 1. *For a feedback channel with capacity B bits per channel use, the Grassmannian codebook as defined in Definition 3 is the same as the set of cluster centroids found by the K -means algorithm with $K = 2^B$ that minimizes $D(Q_{\mathcal{F}})$ for a given distribution of optimal MRT beamforming vector through its training dataset as given in (10), i.e. $\mathcal{F}^* = \mathcal{F}^K$, where \mathcal{F}^K is given by (11).*

Proof: The optimal codebook is given by

$$\mathcal{F}^* = \arg \min_{\substack{\mathcal{F} \subset \mathcal{G}(M_t, 1) \\ |\mathcal{F}| = 2^B}} L_{ub}(\mathcal{F}) = \arg \min_{\substack{\mathcal{F} \subset \mathcal{G}(M_t, 1) \\ |\mathcal{F}| = 2^B}} E_{\mathbf{v}_1} [1 - |\mathbf{v}_1^H \mathbf{f}|^2]$$

$$\begin{aligned} &= \arg \min_{\substack{\mathcal{F} \subset \mathcal{G}(M_t, 1) \\ |\mathcal{F}| = 2^B}} E_{\mathbf{v}_1} \left[\min_{\mathbf{f}_i \in \mathcal{F}} (1 - |\mathbf{v}_1^H \mathbf{f}_i|^2) \right] \\ &= \arg \min_{\substack{\mathcal{F} \subset \mathcal{G}(M_t, 1) \\ |\mathcal{F}| = 2^B}} E_{\mathbf{v}_1} \left[\min_{\mathbf{f}_i \in \mathcal{F}} d^2(\mathbf{v}_1, \mathbf{f}_i) \right] = \mathcal{F}^K, \end{aligned}$$

which completes the proof. \blacksquare

Theorem 1 states the equivalence of the optimal codebook design with the K -means clustering on $\mathcal{G}(M_t, 1)$. The benefit of making this connection is that it provides an approach for finding the optimal codebooks leveraging existing work on K -means clustering on $\mathcal{G}(M_t, 1)$. We are now in a position to state the key steps in designing the optimal codebooks based on the Grassmannian K -means clustering.

A. Codebook Construction

We assume a stationary distribution of the channel for a given coverage area of a transmitter. In order to construct the Grassmannian codebook, we construct $\mathcal{H} = \{\mathbf{H}\}$, a set of channel realizations sampled for different user locations. The available channel dataset \mathcal{H} is split into training and testing datasets, $\mathcal{H}_{\text{train}}$ and $\mathcal{H}_{\text{test}}$ for generating beamforming codebooks and evaluating their performance, respectively. We assume that the size of the training set is large enough so that the sampling distribution closely approximates the original distribution. The training procedure yields the optimal codebook whose performance is evaluated by measuring Γ_{av} for the channel realizations in the test set $\mathcal{H}_{\text{test}}$. Further details and benchmarking results are outlined in Section VI. The codebook design and performance evaluation processes are illustrated in Alg. 2.

Algorithm 2 Training and testing of the Grassmannian codebook

```

1: procedure TRAINCODEBOOK( $\mathcal{H}_{\text{train}}, B$ )
2:   Initialize training set  $\mathcal{X}_{\text{train}} = \emptyset$  on  $\mathcal{G}(M_t, 1)$ 
3:   for  $\mathbf{H} \in \mathcal{H}_{\text{train}}$  do
4:      $\mathbf{U}\Sigma\mathbf{V}^H \leftarrow \text{svd}(\mathbf{H})$ 
5:      $\mathcal{X}_{\text{train}} \leftarrow \mathcal{X}_{\text{train}} \cup \mathbf{v}_1$ 
6:    $\mathcal{F}^* \leftarrow \text{CODEBOOK}(\mathcal{X}_{\text{train}}, K = 2^B)$ 
   return  $\mathcal{F}^*$ 

7: procedure TESTCODEBOOK( $\mathcal{H}_{\text{test}}, \mathcal{F}^*$ )
8:   Initialize  $\Gamma_{av} = 0$ 
9:   for  $\mathbf{H} \in \mathcal{H}_{\text{test}}$  do
10:     $\mathbf{U}\Sigma\mathbf{V}^H \leftarrow \text{svd}(\mathbf{H})$ 
11:     $\mathbf{f} \leftarrow \arg \min_{\mathbf{f}' \in \mathcal{F}^*} d^2(\mathbf{v}_1, \mathbf{f}')$ 
12:     $\Gamma_{av} \leftarrow \Gamma_{av} + \frac{1}{\#\mathcal{H}_{\text{test}}} \left( \frac{\Gamma(\mathbf{f})}{\Gamma(\mathbf{v}_1)} \right)$ 
   return  $\Gamma_{av}$ 

```

V. GRASSMANNIAN PRODUCT CODEBOOK FOR FD-MIMO

After discussing the general notion of the codebook construction for transmit beamforming for an $M_t \times M_r$ MIMO system, we focus our attention to a special case of FD-MIMO communication. We assume that the transmitter is equipped with a UPA with dimensions $M_v \times M_h$ ($M_h M_v = M_t$) while

the receiver has one antenna, i.e. $M_r = 1$. Let $\tilde{\mathbf{H}}$ represent the $M_v \times M_h$ channel matrix where the (i, j) -th element corresponds to the channel between the antenna element at the i -th row and j -th column of the UPA and the single receiver antenna at the user. Note that this system model appears as a special case of the general $M_t \times M_r$ MIMO system discussed in the previous section where $\mathbf{H}^T = \text{vec}(\tilde{\mathbf{H}}^T)$. Hence the codebook can be designed as given in Alg. 2 using the K -means clustering in $\mathcal{G}(M_v M_h, 1)$. Assuming $M_v = M_h = O(n)$, the dimension of the codewords increase as $O(n^2)$. Naturally, the K -means clustering will suffer from the *curse of dimensionality* as the difference in the maximum and minimum distances between two points in the dataset becomes less prominent as the dimension of the space increases [22]. In this section, we show that the codebook can be obtained by clustering on lower dimensional manifolds by exploring the geometry of the UPA. Considering the UPA channel as a matrix $\tilde{\mathbf{H}}$, we have the SVD of $\tilde{\mathbf{H}}$ as follows:

$$\tilde{\mathbf{H}} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H, \quad (17)$$

where \mathbf{U} is the left singular matrix, \mathbf{V} is the right singular matrix and $\mathbf{\Sigma}$ is the rectangular diagonal matrix with singular values σ_i in decreasing order. Let λ_i be the i -th eigenvalue of $\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}$, then $\lambda_i = \sigma_i^2$. Further, \mathbf{u}_i and \mathbf{v}_i are the column vectors of \mathbf{U} and \mathbf{V} respectively with $\mathbf{u}_i^H \mathbf{u}_i = 1$ and $\mathbf{v}_i^H \mathbf{v}_i = 1$. Thus, $\mathbf{u}_i \in \mathcal{U}(M_v, 1)$ and $\mathbf{v}_i \in \mathcal{U}(M_h, 1)$. Then we have

$$\begin{aligned} \mathbf{H}^T &= \text{vec}(\tilde{\mathbf{H}}^T) = \text{vec}(\mathbf{V}^* \mathbf{\Sigma} \mathbf{U}^T) \\ &= \text{vec} \left(\sum_{i=1}^{\text{rank}(\tilde{\mathbf{H}})} \sigma_i \mathbf{v}_i^* \mathbf{u}_i^T \right) = \sum_{i=1}^{\text{rank}(\tilde{\mathbf{H}})} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i^*. \end{aligned} \quad (18)$$

From (18), we can represent \mathbf{H} as the linear combination of $\mathbf{u}_i^T \otimes \mathbf{v}_i^H$ scaled with σ_i . Thus we have $\mathbf{H} = \sum_{i=1}^{\text{rank}(\tilde{\mathbf{H}})} \sigma_i \mathbf{u}_i^T \otimes \mathbf{v}_i^H$.

Due to the finiteness of the physical paths between the transmitter to the receiver, it is well-known that $\text{rank}(\mathbf{H}) \ll \min(M_v, M_h)$. For the sake of simplicity, we approximate the channel \mathbf{H} with its dominant direction, i.e. $\mathbf{u}_1^T \otimes \mathbf{v}_1^H$, which is called rank-1 approximation and the approximated channel $\bar{\mathbf{H}}$ is given as

$$\mathbf{H} \approx \bar{\mathbf{H}} = \sigma_1 \mathbf{u}_1^T \otimes \mathbf{v}_1^H. \quad (19)$$

Let $\mathbf{x} \in \mathcal{U}(M_t, 1)$ be a beamforming vector for $\bar{\mathbf{H}}$. Then the KP form of $\bar{\mathbf{H}}$ naturally leads us to the idea of using \mathbf{x} of the form $\mathbf{x} = \mathbf{x}_v \otimes \mathbf{x}_h$. This motivates us to use separate codebooks \mathcal{F}_h and \mathcal{F}_v for the horizontal and vertical dimensions and enables to design product codebooks by clustering in lower dimensional manifolds. The beamforming gain $\Gamma(\mathbf{x})$ for $\bar{\mathbf{H}}$ can now be simplified as $\Gamma(\mathbf{x}) = \|\bar{\mathbf{H}}\mathbf{x}\|_2^2 =$

$$\begin{aligned} \|\sigma_1 (\mathbf{u}_1^T \otimes \mathbf{v}_1^H) (\mathbf{x}_v \otimes \mathbf{x}_h)\|_2^2 &\stackrel{(a)}{=} \sigma_1^2 \|\mathbf{u}_1^T \mathbf{x}_v\|_2^2 \|\mathbf{v}_1^H \mathbf{x}_h\|_2^2 \\ &= \sigma_1^2 |\mathbf{u}_1^T \mathbf{x}_v|^2 |\mathbf{v}_1^H \mathbf{x}_h|^2, \end{aligned}$$

where step (a) follows from the fact that $\|\mathbf{A} \otimes \mathbf{B}\|_2 = \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$ for two matrices \mathbf{A} and \mathbf{B} of any dimensions. The optimal MRT beamforming vector \mathbf{f} for $\bar{\mathbf{H}}$ can be simplified as $\mathbf{f} = \arg \max_{\mathbf{x} \in \mathcal{U}(M_t, 1)} \Gamma(\mathbf{x}) =$

$$\begin{aligned} \arg \max_{\substack{\mathbf{x}_v \in \mathcal{U}(M_v, 1) \\ \mathbf{x}_h \in \mathcal{U}(M_h, 1)}} |\mathbf{u}_1^T \mathbf{x}_v|^2 |\mathbf{v}_1^H \mathbf{x}_h|^2 &= \arg \max_{\mathbf{x}_v \in \mathcal{U}(M_v, 1)} |\mathbf{u}_1^T \mathbf{x}_v|^2 \\ &\otimes \arg \max_{\mathbf{x}_h \in \mathcal{U}(M_h, 1)} |\mathbf{v}_1^H \mathbf{x}_h|^2 = \mathbf{f}_v \otimes \mathbf{f}_h, \end{aligned} \quad (20)$$

$$\text{where } \mathbf{f}_v = \arg \max_{\mathbf{x}_v \in \mathcal{U}(M_v, 1)} |\mathbf{u}_1^T \mathbf{x}_v|^2, \mathbf{f}_h = \arg \max_{\mathbf{x}_h \in \mathcal{U}(M_h, 1)} |\mathbf{v}_1^H \mathbf{x}_h|^2.$$

Observe that one possible solution for optimal MRT beamformer \mathbf{f} in (20) is given by $\mathbf{f}_v = \mathbf{u}_1^*$ and $\mathbf{f}_h = \mathbf{v}_1$. Following Remark 1, we can argue that $\mathbf{f} = \mathbf{f}_v \otimes \mathbf{f}_h$, where $\mathbf{f}_u \in \mathcal{G}(M_u, 1)$ and $\mathbf{f}_v \in \mathcal{G}(M_h, 1)$.

The loss in average normalized beamforming gain Γ_{av} with the codebook $\mathcal{F} = \mathcal{F}_v \otimes \mathcal{F}_h$ can be bounded as $L(\mathcal{F}) =$

$$\begin{aligned} \mathbb{E}_{\mathbf{u}_1, \mathbf{v}_1} \left[1 - \frac{\Gamma(\mathbf{f}_v \otimes \mathbf{f}_h)}{\Gamma(\mathbf{u}_1^T \otimes \mathbf{v}_1^H)} \right] &= \mathbb{E}_{\mathbf{u}_1, \mathbf{v}_1} \left[1 - |(\mathbf{u}_1^T \otimes \mathbf{v}_1^H)(\mathbf{f}_v \otimes \mathbf{f}_h)|^2 \right] \\ &\leq 2 \mathbb{E}_{\mathbf{u}_1, \mathbf{v}_1} \left[1 - |(\mathbf{u}_1^T \otimes \mathbf{v}_1^H)(\mathbf{f}_v \otimes \mathbf{f}_h)| \right] \\ &\leq 2 \mathbb{E}_{\mathbf{u}_1, \mathbf{v}_1} \left[\min_{\theta, \phi} (|(e^{j\theta} \mathbf{u}_1^* \otimes e^{j\phi} \mathbf{v}_1) - (\mathbf{f}_v \otimes \mathbf{f}_h)|) \right] \\ &\leq 2 \mathbb{E}_{\mathbf{u}_1, \mathbf{v}_1} \left[\min_{\theta, \phi} (\|e^{j\theta} \mathbf{u}_1^*\|_2 \|e^{j\phi} \mathbf{v}_1 - \mathbf{f}_h\|_2 + \right. \\ &\quad \left. \|e^{j\theta} \mathbf{u}_1^* - \mathbf{f}_v\|_2 \|e^{j\phi} \mathbf{f}_h\|_2) \right] \\ &= 2 \mathbb{E}_{\mathbf{u}_1, \mathbf{v}_1} \left[\min_{\theta, \phi} (\|e^{j\phi} \mathbf{v}_1 - \mathbf{f}_h\|_2 + \|e^{j\theta} \mathbf{u}_1^* - \mathbf{f}_v\|_2) \right] \\ &= 2 \mathbb{E}_{\mathbf{u}_1, \mathbf{v}_1} \left[(1 - |\mathbf{v}_1^H \mathbf{f}_h|)^{1/2} + (1 - |\mathbf{u}_1^T \mathbf{f}_v|)^{1/2} \right] \\ &\leq 2 \mathbb{E}_{\mathbf{u}_1, \mathbf{v}_1} \left[(1 - |\mathbf{v}_1^H \mathbf{f}_h|^2)^{1/2} + (1 - |\mathbf{u}_1^T \mathbf{f}_v|^2)^{1/2} \right] = L_{ub}(\mathcal{F}). \end{aligned} \quad (21)$$

Definition 4 (Grassmannian product codebook). *Under the rank-1 approximation of the channel, $\mathbf{H} \approx \bar{\mathbf{H}} = \sigma_1 \mathbf{u}_1^T \otimes \mathbf{v}_1^H$, the B -bit Grassmannian product codebook $\mathcal{F}^* = \mathcal{F}_v^* \otimes \mathcal{F}_h^*$ is the one that satisfies the codebook design criteria in Definition 3 for a given $[B_v, B_h]$ where $|\mathcal{F}_v^*| = 2^{B_v}$, $|\mathcal{F}_h^*| = 2^{B_h}$ and $B_h + B_v = B$.*

We will now state the method to construct the product codebook \mathcal{F}^* as follows.

Lemma 3. *The Grassmannian product codebook $\mathcal{F}^* = \mathcal{F}_v^* \otimes \mathcal{F}_h^*$ as defined in Definition 4 is constructed using the set of centroids \mathcal{F}_v^K and \mathcal{F}_h^K obtained from the independent K -means clustering of the optimal MRT beamforming vectors \mathbf{u}_1^* and \mathbf{v}_1 on $\mathcal{G}(M_v, 1)$ and $\mathcal{G}(M_h, 1)$ with $K = 2^{B_v}$ and $K = 2^{B_h}$, respectively.*

Proof: From Definition 4,

$$\begin{aligned} \mathcal{F}^* &= \mathcal{F}_v^* \otimes \mathcal{F}_h^* = \arg \min_{\mathcal{F}_v, \mathcal{F}_h} L_{ub}(\mathcal{F}) \\ &= \arg \min_{\mathcal{F}_v, \mathcal{F}_h} \mathbb{E}_{\mathbf{u}_1, \mathbf{v}_1} \left[(1 - |\mathbf{v}_1^H \mathbf{f}_h|^2) + (1 - |\mathbf{u}_1^T \mathbf{f}_v|^2) \right] \\ &= \arg \min_{\substack{\mathcal{F}_v \subset \mathcal{G}(M_v, 1) \\ |\mathcal{F}_v| = 2^{B_v}}} \mathbb{E}_{\mathbf{u}_1} (1 - |\mathbf{u}_1^T \mathbf{f}_v|^2) \otimes \\ &\quad \arg \min_{\substack{\mathcal{F}_h \subset \mathcal{G}(M_h, 1) \\ |\mathcal{F}_h| = 2^{B_h}}} \mathbb{E}_{\mathbf{v}_1} (1 - |\mathbf{v}_1^H \mathbf{f}_h|^2), \end{aligned}$$

$$\text{where } \mathcal{F}_v^* = \arg \min_{\substack{\mathcal{F}_v \subset \mathcal{G}(M_v, 1) \\ |\mathcal{F}|=2^{B_v}}} \mathbb{E}_{\mathbf{u}_1} [1 - |\mathbf{u}_1^T \mathbf{f}_v|^2],$$

$$\mathcal{F}_h^* = \arg \min_{\substack{\mathcal{F}_h \subset \mathcal{G}(M_h, 1) \\ |\mathcal{F}|=2^{B_h}}} \mathbb{E}_{\mathbf{v}_1} [1 - |\mathbf{v}_1^H \mathbf{f}_h|^2].$$

Following Theorem 1, we have $\mathcal{F}_h^* = \mathcal{F}_h^K, \mathcal{F}_v^* = \mathcal{F}_v^K$. Thus, the Grassmannian product codebook is given as

$$\mathcal{F}^* = \mathcal{F}_v^* \otimes \mathcal{F}_h^* = \mathcal{F}_v^K \otimes \mathcal{F}_h^K. \quad (22)$$

From Lemma 3, it is possible to perform K -means clustering independently on $\mathcal{G}(M_v, 1)$, $\mathcal{G}(M_h, 1)$ and construct the product codebook with reduced complexity instead of performing K -means clustering on $\mathcal{G}(M_h M_v, 1)$. The training and testing procedure of the proposed product codebook design for a given set of channel realizations \mathcal{H} is given in the following remark.

Remark 3. For a training set $\mathcal{H}_{\text{train}}$, the Grassmann product codebook \mathcal{F}^* defined as $\mathcal{F}^* = \mathcal{F}_v^* \otimes \mathcal{F}_h^*$, where \mathcal{F}_v^* and \mathcal{F}_h^* are obtained by the procedure TRAINPRODCODEBOOK($\mathcal{H}_{\text{train}}, [B_v, B_h]$) using Alg. 3, where B_v, B_h are the number of bits used to encode $\mathbf{u}_1^*, \mathbf{v}_1$ respectively ($B = B_v + B_h$). The performance of the codebook \mathcal{F}^* is evaluated by the procedure TESTPRODCODEBOOK($\mathcal{H}_{\text{test}}, [\mathcal{F}_v^*, \mathcal{F}_h^*]$) as given in Alg. 3.

Algorithm 3 Training and testing of the Grassmannian product codebook

```

1: procedure TRAINPRODCODEBOOK( $\mathcal{H}_{\text{train}}, [B_v, B_h]$ )
2:   Initialize training sets  $\mathcal{X}_{\text{train}} = \emptyset$  and  $\mathcal{Y}_{\text{train}} = \emptyset$  on
    $\mathcal{G}(M_h, 1)$  and  $\mathcal{G}(M_v, 1)$  respectively
3:   for  $\mathbf{H} \in \mathcal{H}_{\text{train}}$  do
4:     Generate  $\tilde{\mathbf{H}}$  from  $\mathbf{H}, \mathbf{U}\Sigma\mathbf{V}^H \leftarrow \text{svd}(\tilde{\mathbf{H}})$ 
5:      $\mathcal{X}_{\text{train}} \leftarrow \mathcal{X}_{\text{train}} \cup \mathbf{v}_1, \mathcal{Y}_{\text{train}} \leftarrow \mathcal{Y}_{\text{train}} \cup \mathbf{u}_1^*$ 
6:    $\mathcal{F}_h^* \leftarrow \text{CODEBOOK}(\mathcal{X}_{\text{train}}, 2^{B_h})$ 
7:    $\mathcal{F}_v^* \leftarrow \text{CODEBOOK}(\mathcal{Y}_{\text{train}}, 2^{B_v})$ 
   return  $[\mathcal{F}_v^*, \mathcal{F}_h^*]$ 

8: procedure TESTPRODCODEBOOK( $\mathcal{H}_{\text{test}}, [\mathcal{F}_v^*, \mathcal{F}_h^*]$ )
9:   Initialize  $\Gamma_{\text{av}} = 0$ 
10:  for  $\mathbf{H} \in \mathcal{H}_{\text{test}}$  do
11:    Generate  $\tilde{\mathbf{H}}$  from  $\mathbf{H}, \mathbf{U}\Sigma\mathbf{V}^H \leftarrow \text{svd}(\tilde{\mathbf{H}})$ 
12:     $\mathbf{f}_h \leftarrow \arg \min_{\mathbf{f}' \in \mathcal{F}_h^*} d^2(\mathbf{v}_1, \mathbf{f}')$ 
13:     $\mathbf{f}_v \leftarrow \arg \min_{\mathbf{f}' \in \mathcal{F}_v^*} d^2(\mathbf{u}_1^*, \mathbf{f}')$ 
14:     $\Gamma_{\text{av}} \leftarrow \Gamma_{\text{av}} + \frac{1}{\#\mathcal{H}_{\text{test}}} \frac{\Gamma(\mathbf{f}_v \otimes \mathbf{f}_h)}{\Gamma(\mathbf{u}_1^* \otimes \mathbf{v}_1)}$ 
  return  $\Gamma_{\text{av}}$ 

```

VI. SIMULATIONS AND DISCUSSIONS

The proposed learning framework requires channel datasets to construct the corresponding beamforming codebooks. In this section, we describe the scenario adopted for the channel dataset generation and evaluate the performance of the generated codebooks in terms of Γ_{av} .

TABLE I: Parameters of the DeepMIMO dataset

Name of scenario	II_2p5
Active BS	3
Active users	1 to 704
Number of antennas (x, y, z)	($M_v, M_h, 1$)
System bandwidth	0.2 GHz
Antennas spacing	0.5
Number of OFDM sub-carriers	1
OFDM sampling factor	1
OFDM limit	1

A. Dataset generation

In this paper, we consider an indoor communication scenario between the base station and the users with $M_r = 1$ operating at a frequency of 2.5 GHz. The described scenario is a part of the DeepMIMO dataset [23]. The channel dataset with the parameters given in Table. I is generated using the dataset generation script of DeepMIMO.

B. Results

We now demonstrate the performance of the proposed codebook design techniques through simulations. The dataset generated as described in Section VI-A is represented as $\mathcal{H} = \{\mathbf{H}\}$ where $\mathbf{H} \in \mathbb{C}^{1 \times M_t}$ and partitioned into training and testing datasets, $\mathcal{H}_{\text{train}}$ and $\mathcal{H}_{\text{test}}$ respectively. We refer to the Grassmannian codebook design in Section IV as Prop. I: $[B]$ when B bits are used for generating the codebook as illustrated in Alg. 2. The Grassmannian product codebook design described in Section V is referred as Prop. II: $[B_v, B_h]$ where $[B_v, B_h]$ is the allocation of the feedback bits for constructing the codebooks using Alg. 3 and $B = B_v + B_h$.

We compare the average normalized beamforming gain Γ_{av} of the two proposed codebooks with that of the DFT structured KP codebooks [24] (referred to as KP DFT) and the codebooks generated based on the Grassmannian line packings for correlated channel [8] (referred to as GLP). The Grassmannian line packings required for the codebook construction in [8] were obtained from [7], [25]. The channel correlation matrix \mathbf{R} is calculated from the training channel dataset $\mathcal{H}_{\text{train}}$ according to $\mathbf{R} = \mathbb{E}_{\mathbf{H}}(\mathbf{H}^H \mathbf{H})$.

In Fig. 4, the normalized beamforming gains for different transmit antenna configurations $M_v \times M_h$ and feedback-bit allocations $[B_v, B_h]$ using the four codebook design techniques are plotted. The KP codebooks relying on DFT structure are simple to construct but the quantized beams may not be effective because the KP codebooks search only the beams lying in the direction of the right and left dominant singular vectors of the reshaped channel in (17). While the Grassmannian line packings were shown to be the optimal codebooks for uncorrelated Rayleigh fading channel and were extended to correlated Rayleigh fading channel, they do not necessarily work well with any arbitrary channel distribution. It was not possible to show the performance of the GLP codebooks for all $M_v \times M_h$ because of the challenges in finding the best line packings. Simulation results in Fig. 4 demonstrate that the codebooks designed using the proposed techniques adapt well to the underlying channel distribution. They also outperform KP DFT codebooks [24], GLP codebooks [8] and

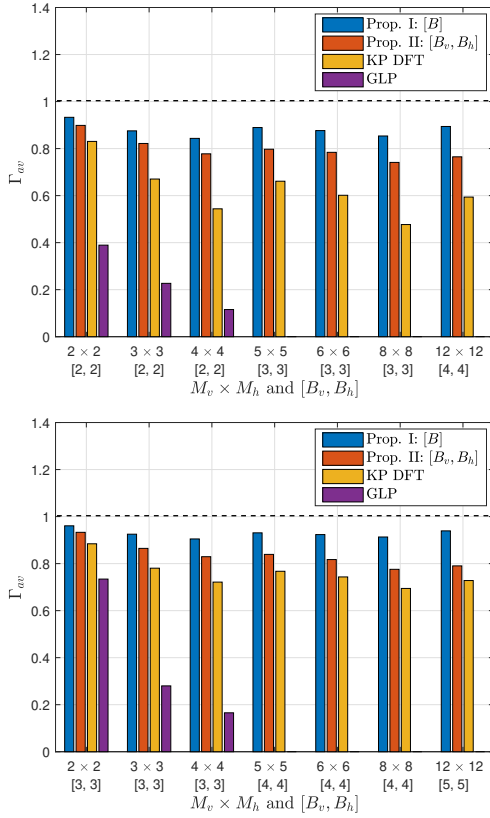


Fig. 4: Average normalized beamforming gain for different transmit antenna configurations $M_v \times M_h$ and feedback-bit allocations $[B_v, B_h]$.

provide beamforming gains comparable to that of optimal MRT beamforming. Equivalently, they reduce the feedback overhead because they can maintain the same quantization performance with less overhead compared to the other codebooks.

VII. CONCLUSION

This paper considered the problem of codebook based MRT beamforming in an FDD-MIMO system operating under arbitrary channel conditions. Leveraging well-known connections of this problem with Grassmannian line packing, we identified that the problem of finding the optimal MRT beamforming codebooks for any arbitrary channel distribution can be constructed using the K -means clustering of MRT beamforming vectors on $\mathcal{G}(M_t, 1)$. We presented the Grassmannian K -means clustering algorithm to construct the beamforming codebooks. We showed that this approach can be extended to FD-MIMO systems with UPA antennas to design product codebooks with reduced computational complexity. As our future work, we will be investigating the design of optimal codebooks for limited feedback unitary precoding for spatial multiplexing in MIMO systems with arbitrary underlying channel. Overall, this paper provides a concrete example of a design problem with rigorous mathematical underpinnings that benefits from a classical *shallow* learning approach.

REFERENCES

[1] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Journal on Sel. Areas in Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct 1998.

[2] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Sel. Areas in Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct 2008.

[3] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. on Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct 2003.

[4] P. Xia and G. B. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. on Signal Proc.*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[5] K. Huang, R. W. Heath, and J. G. Andrews, "Limited feedback beamforming over temporally-correlated channels," *IEEE Trans. on Signal Proc.*, vol. 57, no. 5, pp. 1959–1975, May 2009.

[6] B. Lee, J. Choi, J. Seol, D. J. Love, and B. Shim, "Antenna grouping based feedback compression for FDD-based massive MIMO systems," *IEEE Trans. on Commun.*, vol. 63, no. 9, pp. 3261–3274, Sep 2015.

[7] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. on Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct 2003.

[8] D. J. Love and R. W. Heath, "Limited feedback diversity techniques for correlated channels," *IEEE Trans. on Vehicular Tech.*, vol. 55, no. 2, pp. 718–722, Mar 2006.

[9] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb 2002.

[10] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. on Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug 2005.

[11] Y. Huang, P. P. Liang, Q. Zhang, and Y. Liang, "A machine learning approach to MIMO communications," in *Proc. IEEE ICC*, 2018, pp. 1–6.

[12] Y. Du, G. Zhu, J. Zhang, and K. Huang, "Automatic recognition of space-time constellations by learning on the Grassmann manifold," *IEEE Trans. on Signal Proc.*, vol. 66, no. 22, pp. 6031–6046, Nov 2018.

[13] A. Abdelnasser, E. Hossain, and D. I. Kim, "Clustering and resource allocation for dense femtocells in a two-tier cellular OFDMA network," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 3, pp. 1628–1641, Mar 2014.

[14] J. B. Andersen, "Antenna arrays in mobile communications: gain, diversity, and channel capacity," *IEEE Antennas and Propagation Mag.*, vol. 42, no. 2, pp. 12–16, Apr 2000.

[15] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. on Commun.*, vol. 47, no. 10, pp. 1458–1461, Oct 1999.

[16] C. H. Tse, K. W. Yip, and T. S. Ng, "Performance tradeoffs between maximum ratio transmission and switched-transmit diversity," in *Proc. IEEE PIMRC*, vol. 2, 2000, pp. 1485–1489.

[17] W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*. Academic press, 1986.

[18] J. Zhang, G. Zhu, R. W. Heath, and K. Huang, "Grassmannian learning: Embedding geometry awareness in shallow and deep learning," *arXiv preprint arXiv:1808.02229*, 2018.

[19] J. H. Conway, R. H. Hardin, and N. J. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experimental mathematics*, vol. 5, no. 2, pp. 139–159, Apr 1996.

[20] P. Dixon, A. H. El-Shaarawi, and W. W. Piegorsch, "Ripley's K function," *Encyclopedia of Environmetrics*, vol. 3, 2012.

[21] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. on Commun.*, vol. 28, no. 1, pp. 84–95, Jan 1980.

[22] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2009.

[23] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. ITA*, Feb 2019, pp. 1–8.

[24] J. Choi, K. Lee, D. J. Love, T. Kim, and R. W. Heath, "Advanced limited feedback designs for FD-MIMO using uniform planar arrays," in *Proc. IEEE GLOBECOM*, 2015, pp. 1–6.

[25] A. Medra and T. N. Davidson, "Flexible codebook design for limited feedback systems via sequential smooth optimization on the Grassmannian manifold," *IEEE Trans. on Signal Proc.*, vol. 62, no. 5, pp. 1305–1318, Mar 2014.