Lecture 6: Case Study: Role of Comm Theory in ML: Grassmann Clustering in Massive MIMO

Harpreet S. Dhillon

Wireless@VT, Bradley Department of Electrical & Computer Engineering
Virginia Tech, Blacksburg, VA

https://www.dhillon.ece.vt.edu
hdhillon@vt.edu

Joint work with Keerthana Bhogi and Chiranjib Saha

JTG/IEEE Information Theory Society Summer School
IIT Kanpur

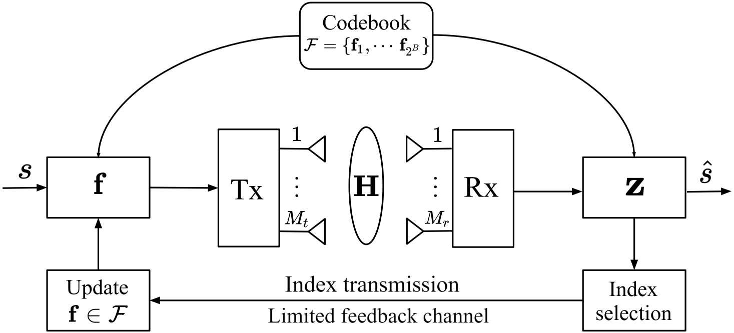
References

- References for this specific case study:
  - K. Bhogi, C. Saha, H. S. Dhillon, "Learning on a Grassmann Manifold: CSI Quantization for Massive MIMO Systems", in *Proc. Asilomar*, Nov. 2020.
  - K. Bhogi, C. Saha, H. S. Dhillon, "Tensor Learning-based Precoder Codebooks for FD-MIMO Systems", arXiv:2106.11374.
- Some Classical References:
  - 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 J. on Sel. Areas in Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct 1998.
  - 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.
  - 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.

Codebook-based Beamforming: Motivation

- Transmit beamforming is one of the simplest approaches to achieve full diversity in MIMO systems.
- Requires channel state information (CSI) at the Tx in the form of transmit beamforming vector.
- Challenging in FDD large-scale MIMO systems that cannot utilize channel reciprocity to acquire CSI.
- Results in significant feedback overhead when the number of antennas is large.
- One solution: Construct a set of beamforming vectors constituting a codebook, which is known to both the Tx and the Rx.
- The problem reduces to determining the best beamforming vector at the Rx and conveying its index to the Tx over the feedback channel.
- Question: What is the best way to construct these codebooks?

Block Diagram and Notation



- We consider M_t transmit antennas and $M_r = 1$ receive antennas.
- $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ is the block fading MIMO channel.
- $s \in \mathbb{C}$: transmitted data symbol.
- $\mathbf{f} \in \mathbb{C}^{M_t \times 1}$, $\mathbf{z} \in \mathbb{C}^{M_r \times 1}$ are the Tx beamforming and Rx combining vectors.
- $\mathcal{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_{2^B}\}$ is the codebook, where B bits per channel use is the capacity of the error-free feedback channel.

Problem Formulation: Basic Setting

- Assuming $\mathbf{n} \sim \mathcal{CN}(0, N_o \mathbf{I}_{M_r})$ to be additive white Gaussian noise, we have

$$\mathbf{y} = \mathbf{H}\mathbf{f}s + \mathbf{n}$$
$$\hat{s} = \mathbf{z}^H \mathbf{H}\mathbf{f}s + \mathbf{z}^H \mathbf{n}$$

- We assume perfect channel knowledge at the receiver, which means it can determine optimal beamforming and combining vectors.
- Total transmit power: $\mathbb{E}[\|\mathbf{f}s\|_2^2] = \mathbb{E}[|s|^2]\,\|\mathbf{f}\|_2^2 = \mathcal{E}_t\,\|\mathbf{f}\|_2^2$.
- For the receiver combining vector, we have $\|\mathbf{z}\|_2^2 = 1$, for which the receiver SNR γ_r is

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

where $\gamma_t = \mathcal{E}_t \|\mathbf{f}\|^2 / N_o$ is the transmit SNR and $\Gamma(\mathbf{f}, \mathbf{z})$ is the beamforming gain. Let's look at the beamforming gain carefully now.

Problem Formulation: Max Beamforming Gain for a given $\mathbf{f}$

- Recall from the previous slide that the beamforming gain is defined as

$$\Gamma(\mathbf{f}, \mathbf{z}) = \frac{|\mathbf{z}^H \mathbf{H}\mathbf{f}|^2}{\|\mathbf{f}\|_2^2}.$$

- Because of the transmit power constraint, it is standard to assume $\|\mathbf{f}\|_2 = 1$, which gives $\Gamma(\mathbf{f}, \mathbf{z}) = |\mathbf{z}^H \mathbf{H}\mathbf{f}|^2$.
- SNR can be maximized at the receiver with $\mathbf{z} = \mathbf{H}\mathbf{f} / \|\mathbf{H}\mathbf{f}\|_2$, which gives $\Gamma(\mathbf{f}) = \|\mathbf{H}\mathbf{f}\|_2^2$ as the maximum beamforming gain for a given $\mathbf{f}$.

Determining Optimal $\mathbf{f}$

Optimization problem: Choose $\mathbf{f}$ such that Γ is maximized.

$$\mathbf{f} = \arg\max_{\mathbf{x}\in\mathbb{C}^{M_t\times 1}} \Gamma(\mathbf{x})$$
$$\text{subject to } \|\mathbf{x}\|_2^2 = 1$$

Observation: If vector $\mathbf{f}$ maximizes $\Gamma(\mathbf{f}) = \|\mathbf{H}\mathbf{f}\|_2^2$, then $\mathbf{f}e^{j\theta}$ for any $\theta \in [0, 2\pi)$ will also maximize $\Gamma(\mathbf{x})$. Therefore, the solution to the above problem is not unique.

A solution to the above problem is $\mathbf{f} = \mathbf{v}_1$, where $\mathbf{v}_1$ is the dominant eigenvector of $\mathbf{H}^H\mathbf{H}$. The maximum value of the beamforming gain is $\Gamma(\mathbf{v}_1) = \|\mathbf{H}\mathbf{v}_1\|_2^2 = \lambda_1$, where λ_1 is the largest eigenvalue of $\mathbf{H}^H\mathbf{H}$.

How is this connected to codebook construction?

Codebook Construction: Intuition

- Let's first assume for a moment that each $\mathbf{H}$ provides a unique optimal $\mathbf{f}$. *We will see shortly how to make this happen.*
  - If we endow a distribution on $\mathbf{H}$, we also get a corresponding distribution on $\mathbf{f}$.
  - Now the problem of codebook construction reduces to determining $\mathcal{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_{2^B}\}$ such that distortion introduced because of selecting a component from $\mathcal{F}$ instead of the actual beamforming vector $\mathbf{f}$ is minimized.
- Challenge: Since we do not get a unique $\mathbf{f}$ for $\mathbf{H}$, the correct way to think about this problem is to think of the codebook as a finite set of *subspaces* in the Euclidean space rather than a finite set of *vectors*. Eventually, we need a distortion measure that is invariant to the phase shifts of $\mathbf{f}$. This can be formally done by posing this problem on the Grassmannian manifold.

Grassmann Manifold: Definition and Context

Definition 1 (Complex Grassmann Manifold $\mathcal{G}(M_t, M)$ [5])

Space formed by *all* M dimensional linear subspaces embedded in $\mathbb{C}^{M_t}$:

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

- $\mathcal{Y} \in \mathcal{G}(M_t, M)$ can be interpreted as a *point* on $\mathcal{G}(M_t, M)$ or a subspace in $\mathbb{C}^{M_t \times 1}$.
- Let $\mathbf{Y}$ be the orthonormal basis that spans $\mathcal{Y}$. Then then $\mathbf{Y}' = \mathbf{Y}\mathbf{R}$ also spans $\mathcal{Y}$ for any unitary matrix $\mathbf{R}$. Therefore, $\mathbf{Y}$ and $\mathbf{Y}'$ map to the same point $\mathcal{Y}$ on the manifold (and we say $\mathbf{Y} \equiv \mathbf{Y}'$).

The above equivalence relation will ensure that the optimal $\mathbf{f}$ for our beam-forming problem is unique on the Grassmann manifold.

Grassmann Manifold: Application to the Beamforming Problem

- Since we are interested in "vectors", the Grassmann manifold of interest for us is $\mathcal{G}(M_t, 1)$. It is just a set of all lines passing through origin in $\mathbb{C}^{M_t \times 1}$.
- 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.
- Let $\mathbf{f}' = \mathbf{f} e^{j\theta}$. It is easy to argue that $\mathbf{f}' \equiv \mathbf{f}$, i.e., they both span the same line $\mathcal{L}$ and hence correspond to the same point on $\mathcal{G}(M_t, 1)$.
  - This is really what we wanted.
- One way to define distance d between two points of $\mathcal{G}(M_t, 1)$ corresponding to $\mathbf{f}_1$ and $\mathbf{f}_2$, respectively, is in terms of the sine of angle between the lines: $d(\mathbf{f}_1, \mathbf{f}_2) = \sin(\theta_{1,2}) = \sqrt{1 - |\mathbf{f}_1^H \mathbf{f}_2|^2}$.
  - It is easy to check that the above distance metric is invariant to phase changes in $\mathbf{f}_1$ or $\mathbf{f}_2$.
  - We will also show that this choice of distance metric is *optimal*.

Clustering on a Grassmann Manifold

- Summary of key takeaways thus far:
  - We have a unique solution $\mathbf{f}$ on the Grassmann manifold for every $\mathbf{H}$.
  - We know how to measure distance between points lying on the manifold.
  - Any probability distribution on channel $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ will impose a probability distribution on $\mathbf{f}$ in $\mathcal{G}(M_t, 1)$.
- Rayleigh fading channels correspond to the uniform distribution on $\mathcal{G}(M_t, 1)$, which reduces the problem of codebook construction to the well-known Grassmann line packing problem.
- However, real-world channels will mostly exhibit clustering, which can be exploited to construct codebooks that are cognizant of the underlying channel distribution.
- This reduces the problem of codebook construction for general channels to the problem of clustering on Grassmann Manifold.

Ripley's K Function: Channel Clustering

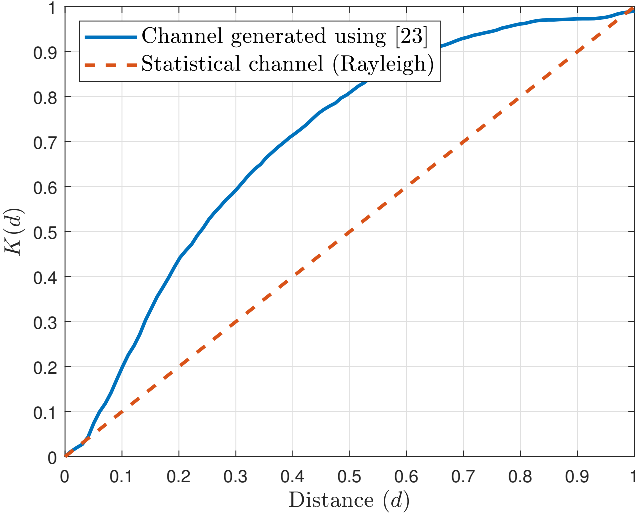


Figure illustrating clustering behaviour of real-world channel with $M_t = 2$, $M_r = 1$ using Ripley's K function.

Grassmannian K-means Clustering

- The idea on the previous slide can now be implemented using K-means clustering on the Grassmann manifold.
- We need to choose K centroids such that the average distortion due to the quantization according to a pre-defined distortion measure is minimized.
- The distortion measure and quantizer are defined as follows.

Definition 2 (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 3 (Grassmann quantizer)

Let $\mathcal{F} \subset \mathcal{G}(M_t, 1)$ be a B-bit codebook such that $\mathcal{F} = \{\mathbf{f}_1, ..., \mathbf{f}_{2^B}\}$, 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. $Q_{\mathcal{F}} : \mathcal{G}(M_t, 1) \mapsto \mathcal{F}$.

Grassmannian K-means Clustering

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

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

  where $\mathbb{E}_{\mathbf{x}}$ means averaging over the dataset $\mathcal{X} = \{\mathbf{x}\}$ in lieu of the probability distribution $p(\mathbf{x})$.
- The set of K centroids $(K = 2^B)$ that minimize $D(Q_{\mathcal{F}})$ is

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

- The associated quantizer is

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

Grassmannian K-means Clustering

- We use Linde-Buzo-Gray (LBG) for K-means clustering.
- The only non-trivial step is the centroid calculation for a set of points.
- The centroid of n elements in a general manifold with respect to an arbitrary distortion measure does not necessarily exist in a closed form.
- Fortunately, the centroid computation on $\mathcal{G}(M_t, 1)$ is feasible because of the following Lemma, which enables extremely efficient construction of the codebooks.

Lemma 1 (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

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

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

Optimality of the Codebook

- In order to define the *optimality* of the codebook, we use the average normalized beamforming gain for $\mathcal{F}$

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

- To measure the average distortion due to quantization, we use the loss in Γ_{av} as given below:

$$L(\mathcal{F}) := \mathbb{E}_{\mathbf{H}}\left[1-\Gamma_{av}\right] \overset{M_r=1}{=} \mathbb{E}_{\mathbf{v}_1}\left[1-|\mathbf{v}_1^H\mathbf{f}|^2\right]. \qquad (1)$$

- Using this we can define the codebook design criterion on the next slide.

Grassmannian Codebook Design

Definition 4 (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 $L(\mathcal{F})$. Therefore $\mathcal{F}^* := \underset{\substack{\mathcal{F}\subset\mathcal{G}(M_t,1)\\ |\mathcal{F}|=2^B}}{\arg\min}\ L(\mathcal{F})$.

Theorem 2

For a feedback channel with capacity B bits per channel use, the Grassmannian codebook as defined in Definition 4 is the same as the set of cluster centroids found by the K-means algorithm for the setting described earlier, i.e. $\mathcal{F}^* = \mathcal{F}^K$.

Grassmannian Product Codebook Design

- Consider full-dimension MIMO communication, where we have a Tx equipped with a UPA with dimensions $M_v \times M_h$ ($M_h M_v = M_t$) while the Rx has one antenna, i.e. $M_r = 1$.
- The codebook can be designed using K-means clustering in $\mathcal{G}(M_v M_h, 1)$. However, K-means clustering may suffer from the well-known *curse of dimensionality* when the number of antennas is high.
- In this work, we proposed a new product codebook design in which we construct the codebook using clustering on lower dimensional manifolds by exploring the geometry of the UPA.
- Unfortunately, we do not have time to cover this in detail but please check the arXiv preprint (arXiv:2106.11374) and email us if you have questions.

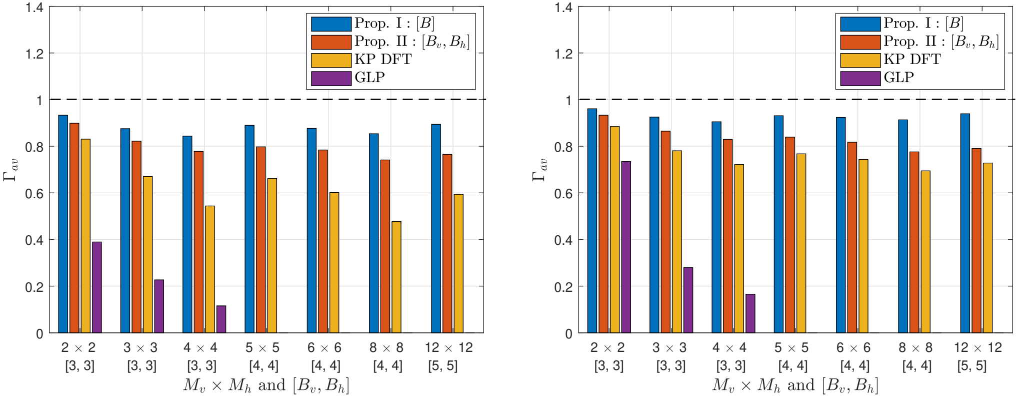
Dataset

- We adopted 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 [4].
- The parameters of the channel dataset are given in Table. 1.

Table: Parameters of the DeepMIMO dataset

Name of scenario	I1_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

Simulation Results



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

Results and Conclusion

- We identified that the optimal beamforming codebooks for any arbitrary channel distribution can be constructed using the K-means clustering of beamforming vectors on $\mathcal{G}(M_t, 1)$.
- These codebooks outperformed both the baselines, namely, Kronecker-product DFT codebooks and Grassmann line packing-based codebooks (even after incorporating correlation using [3]), and provide gains comparable to that of optimal MRT beamforming.
- Since this problem is inspired by the large dimensionality of the channel matrices, the natural tendency is to think in terms of obtaining a lower dimensional representation of the channel using deep learning techniques. However, in this specific problem, we have shown that the analytical structure of the codebook design problem can be leveraged to develop a far more efficient shallow learning technique.

A Useful Resource for ML in Communications

If you are interested in browsing some good papers on machine learning in communications, please check IEEE ComSoc "Best Readings in Machine Learning in Communications". It includes papers on the following topics:

- Signal detection
- Channel encoding and decoding
- Channel estimation, prediction, and compression
- End-to-end communications
- Resource allocation
- Selected topics

The list also includes some overview and tutorial papers.

References

[1] J. Choi *et al.*, "Advanced Limited Feedback Designs for FD-MIMO Using Uniform Planar Arrays," 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, CA, 2015, pp. 1-6.

[2] David J. Love *et al.*, “Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems,“ in IEEE Trans. on Info Theory, vol 49, no. 10, pp. 2735-2747, 2003.

[3] David J. Love *et al.*, “Grassmannian Beamforming on Correlated MIMO Channels,“ in GLOBECOM ’04.

[4] Ahmed Alkhateeb, “Deep MIMO: A Generic Deep Learning Dataset for Millimeter Wave and Massive MIMO Applications,“ in Proc. of Information Theory and Applications Workshop (ITA), Feb 2019

[5] Jiayao Zhang *et al.*, “Grassmannian Learning: Embedding Geometry Awareness in Shallow and Deep Learning”, arXiv:1808.02229.

[6] K. Bhogi *et al.*, “Learning on a Grassmann Manifold: CSI Quantization for Massive MIMO Systems”, in *Proc. Asilomar*, Nov. 2020.