



# GrassCaré

Visualizing **Grassmannians** via Poincaré Embeddings

- Huanran Li, Daniel Pimentel-Alarcón



# What is a Grassmannian?

## Grassmannian

🌐 12 languages ▾

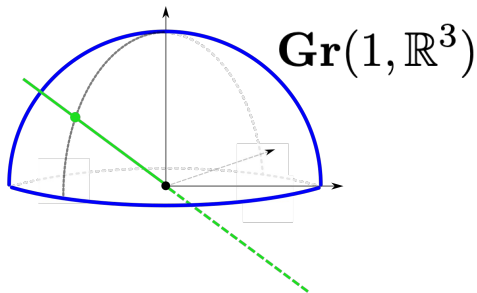
[Article](#) [Talk](#)

[Read](#) [Edit](#) [View history](#)

From Wikipedia, the free encyclopedia

In [mathematics](#), the **Grassmannian**  $\mathbf{Gr}(k, V)$  is a space that parameterizes all  $k$ -dimensional linear subspaces of the  $n$ -dimensional vector space  $V$ . For example, the Grassmannian  $\mathbf{Gr}(1, V)$  is the space of lines through the origin in  $V$ , so it is the same as the [projective space](#) of one dimension lower than  $V$ .<sup>[1][2]</sup>

**Grassmannian**: A space of linear subspaces.





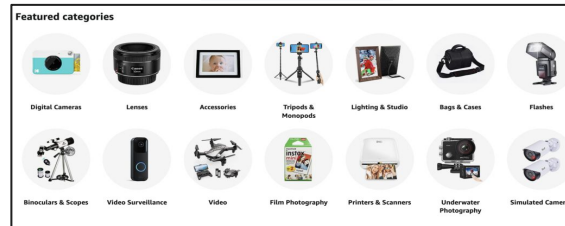
## Recommender System

Type of Users => Subspace

NETFLIX

		Items					
Users							
		10	-1	8	10	9	4
		8	9	10	-1	-1	8
		10	5	4	9	-1	-1
		9	10	-1	-1	-1	3
		6	-1	-1	-1	8	10

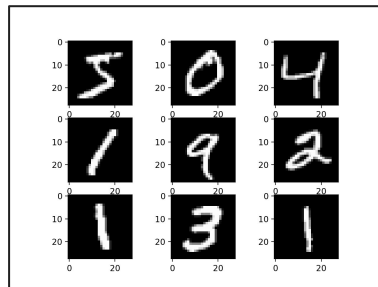
amazon



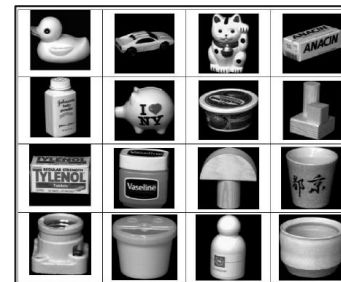
## Object Classification

Class of Image => Subspace

MNIST

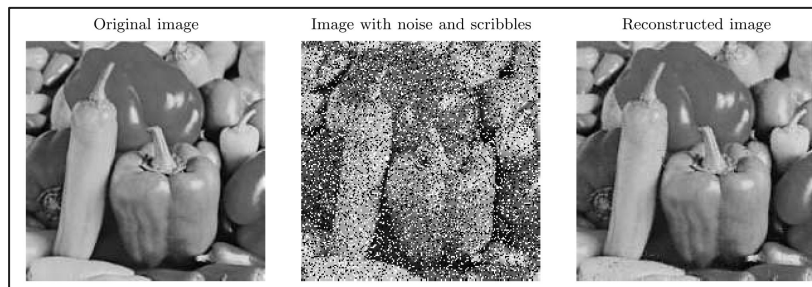


COIL-20



## Image reconstruction

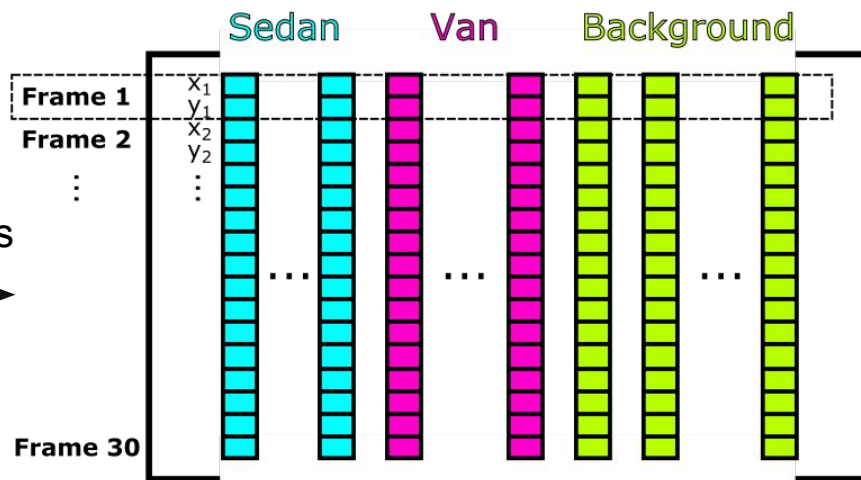
Color Scheme => Subspace



# Object Trajectory



Trajectories



Sample's Pattern



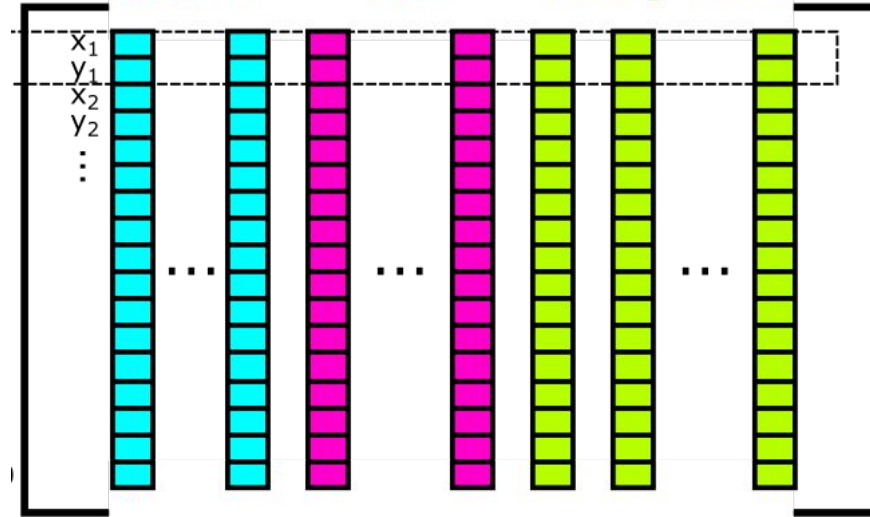
Subspace 1



Subspace 2



Subspace 3



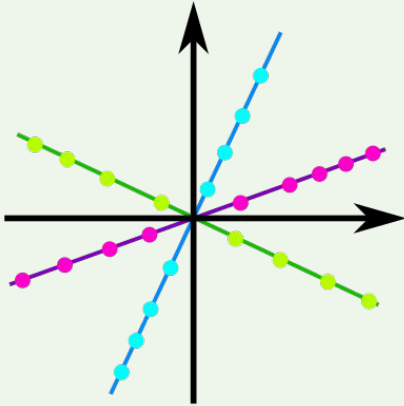
How **many** subspaces?

How **similar** they are?

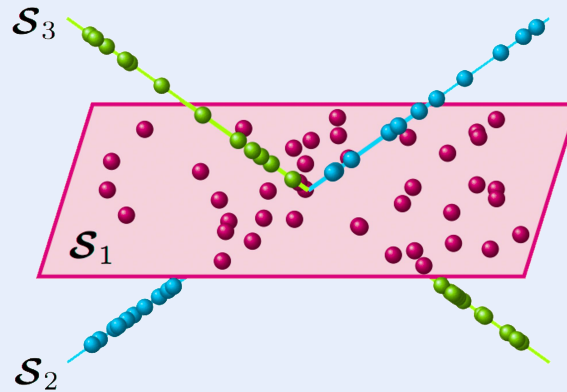
How **many samples** they have?

# Motivation: Visualize Similarity between Subspace

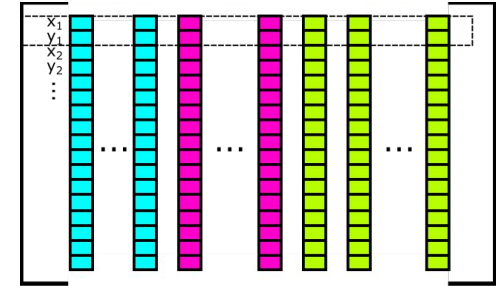
In 2-d Space



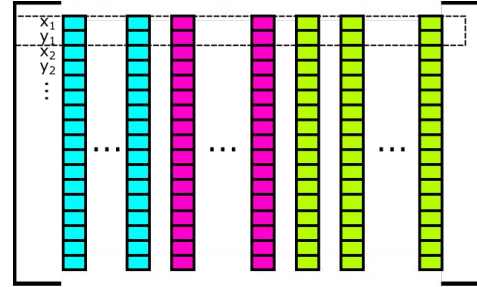
In 3-d Space



In 30-d Space



# How can we visualize High-dimensional subspaces?



**MATRIX**

**SUBSPACE**

**MATRIX**

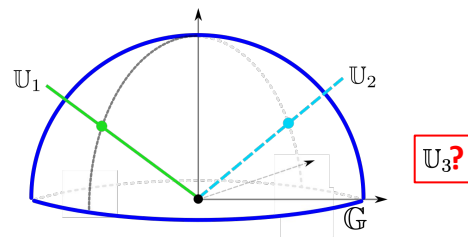
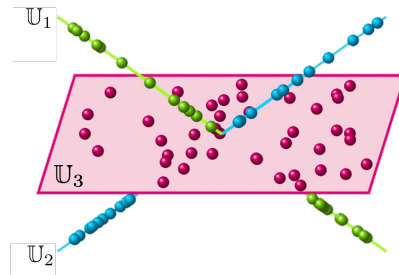
**I SEE A PATTERN, BUT MY IMAGINATION CANNOT PICTURE THE MAKER OF THAT PATTERN.**

**-- ALBERT EINSTEIN IN 1915**

**2023**

# Existing options: Hemisphere Embedding

- **Given:** 1-d subspaces in 3-d space.
- **Output:** Point embedding on 3-d hemisphere.



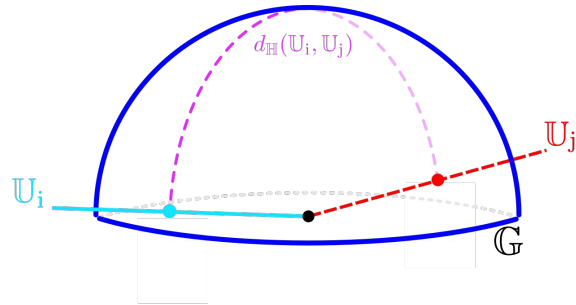
**Drawback #1:** limited to 1-d subspaces (aka lines).  
Very restrictive!



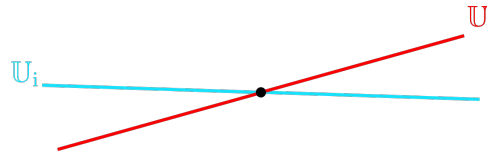
# Existing options: Hemisphere Embedding

**Drawback #2:** Distance can be misleading!

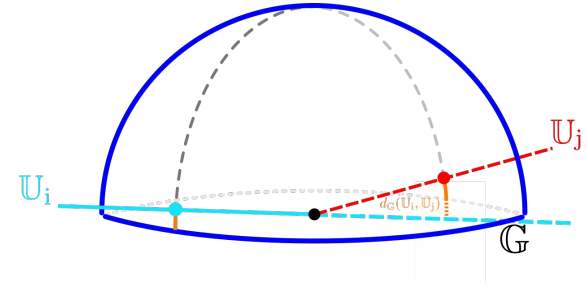
Subspaces Input



Visual Intuition



True distance

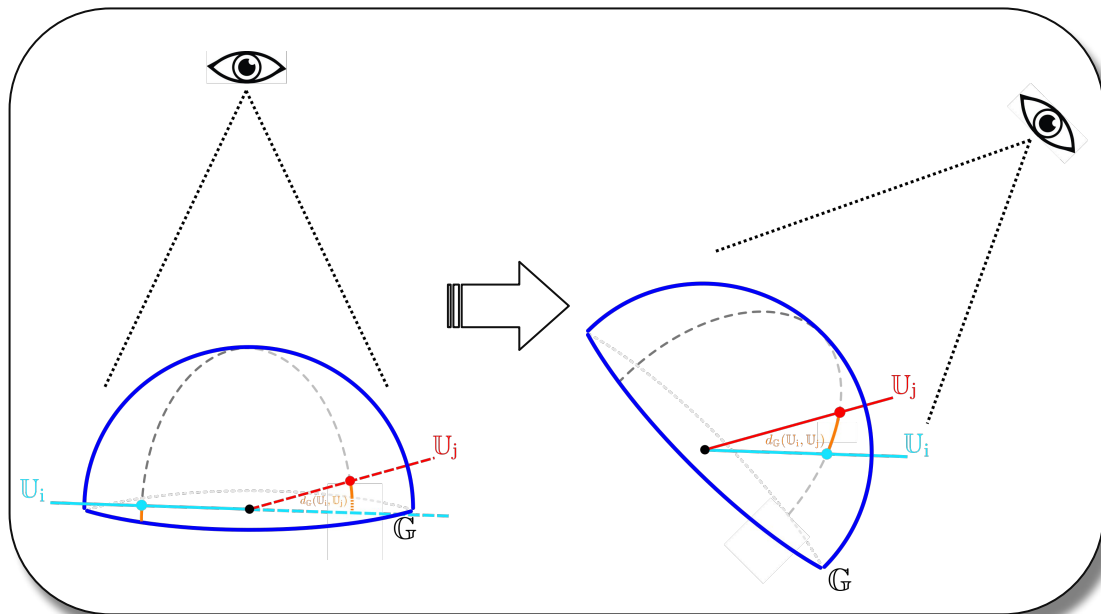


Distance between Subspaces:

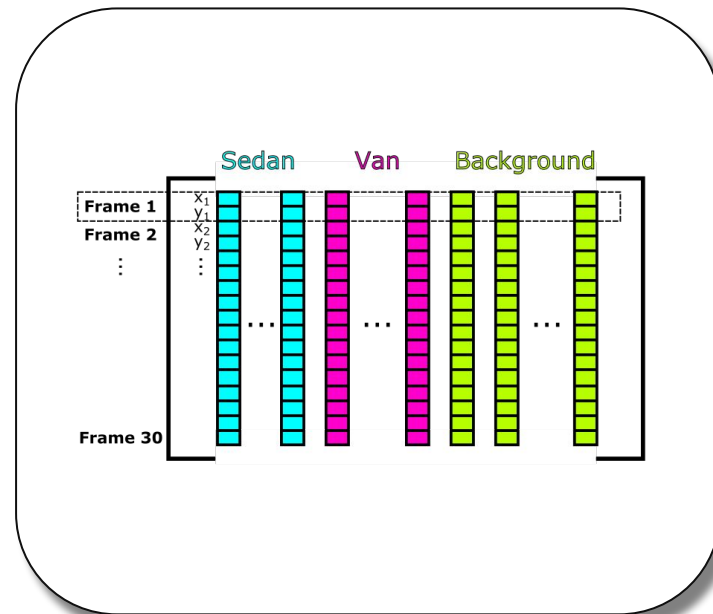
$$d_G(U_i, U_j) := \sqrt{\sum_{\ell=1}^r \arccos^2 \sigma_\ell(U_i^T U_j)}$$

Basis of the subspaces  
l'th largest singular value  
2-Norm of the principle angle

# Goal

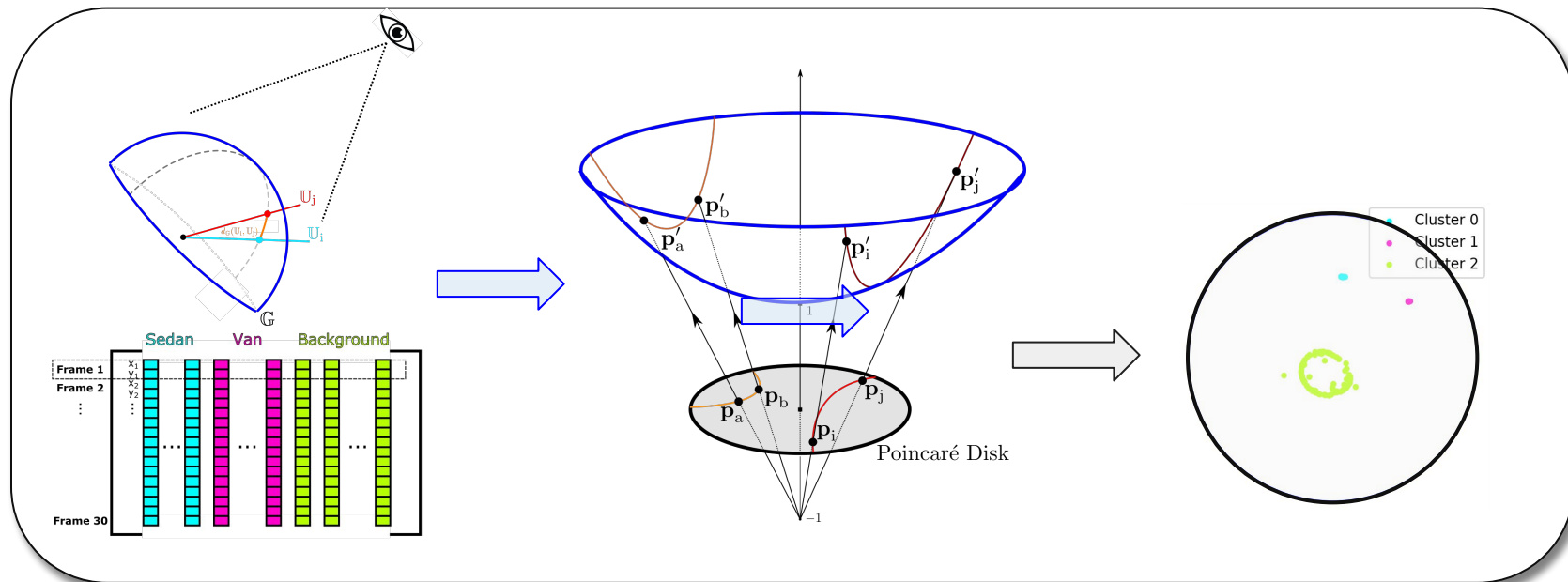


1) Find the right angle



2) high-dimensional subspaces

# How do we find such embedding?





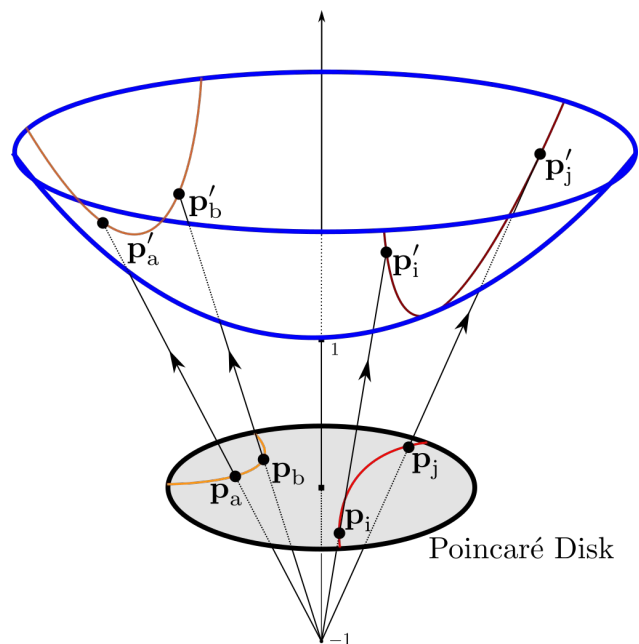
## Distance between points in Poincaré Disk:

$$d_{\mathbb{D}}(\mathbf{p}_i, \mathbf{p}_j) := \operatorname{arcosh} \left( 1 + 2 \frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{(1 - \|\mathbf{p}_i\|^2)(1 - \|\mathbf{p}_j\|^2)} \right)$$

Euclidean Distance

Scaled by how far from origin

**Distances near the edge are larger than they appear.**





# Problem Formulation

- Given: A set of subspace  $\{\mathbb{U}_1, \mathbb{U}_2, \dots, \mathbb{U}_N\}$ .

$$d_G(\mathbb{U}_i, \mathbb{U}_j) := \sqrt{\sum_{\ell=1}^r \arccos^2 \sigma_{\ell}(\mathbf{U}_i^T \mathbf{U}_j)}$$

- Output: A 2-d embeddings  $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$  on the Poincare Disk.

$$d_D(\mathbf{p}_i, \mathbf{p}_j) := \operatorname{arcosh} \left( 1 + 2 \frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{(1 - \|\mathbf{p}_i\|^2)(1 - \|\mathbf{p}_j\|^2)} \right)$$

- Objective: The embedding preserves the most information.

$$\min_{\mathbf{p}_1, \dots, \mathbf{p}_N} \operatorname{KL}(\boxed{\mathbf{P}_G} \parallel \boxed{\mathbf{P}_D})$$

Probability Matrix in Grassmannian

Probability Matrix in Poincaré disk



Probability that  $\mathbb{U}_i$  choose  $\mathbb{U}_j$  as the **closest neighbor**.

$$[\mathbf{P}_G]_{ij} := \frac{1}{2N} \frac{\exp(-d_G(\mathbb{U}_i, \mathbb{U}_j)^2/2\gamma_i^2)}{\sum_{k \neq i} \exp(-d_G(\mathbb{U}_i, \mathbb{U}_k)^2/2\gamma_i^2)} + \frac{1}{2N} \frac{\exp(-d_G(\mathbb{U}_j, \mathbb{U}_i)^2/2\gamma_j^2)}{\sum_{k \neq j} \exp(-d_G(\mathbb{U}_j, \mathbb{U}_k)^2/2\gamma_j^2)} \in \mathbb{R}^{N \times N}$$

Softmin function   Subspace distance

For symmetric purpose

$$\min_{\mathbf{p}_1, \dots, \mathbf{p}_N} \text{KL}(\mathbf{P}_G || \mathbf{P}_D)$$

Softmin function   Poincaré Disk distance

$$[\mathbf{P}_D]_{ij} := \frac{\exp(-d_D(\mathbf{p}_i, \mathbf{p}_j)^2/\beta)}{\sum_{k \neq l} \exp(-d_D(\mathbf{p}_k, \mathbf{p}_l)^2/\beta)} \in \mathbb{R}^{N \times N}$$

Probability that  $\mathbf{p}_i$  choose  $\mathbf{p}_j$  as the **closest neighbor**.



# Riemannian Gradient Descent

**Gradient Descent Step:**  $\mathbf{p}_i^{t+1} \leftarrow R(\mathbf{p}_i^t - \eta \nabla_i \mathcal{L}),$

**Gradient** respect to Poincaré Distance:  $\nabla_i d_{\mathbb{D}}(\mathbf{p}_i, \mathbf{p}_j) = \frac{4}{b\sqrt{c^2 - 1}} \left( \frac{\|\mathbf{p}_j\|^2 - 2\langle \mathbf{p}_i, \mathbf{p}_j \rangle + 1}{a^2} \mathbf{p}_i - \frac{\mathbf{p}_j}{a} \right).$

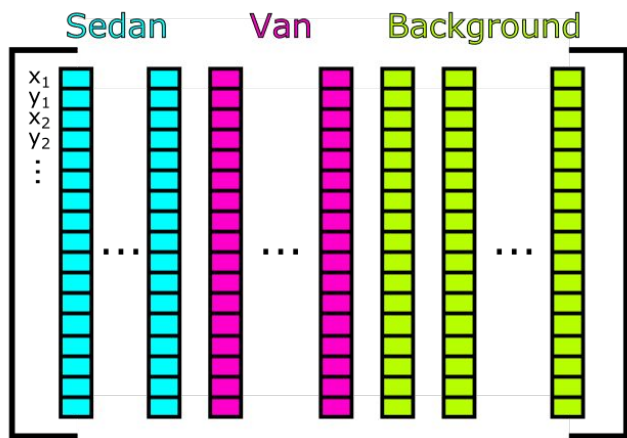
**Gradient** respect to Loss:  $\nabla_i \mathcal{L} = \frac{4}{\beta} \sum_j ([\mathbf{P}_{\mathbb{G}}]_{ij} - [\mathbf{P}_{\mathbb{D}}]_{ij})(1 + d_{\mathbb{D}}(\mathbf{p}_i, \mathbf{p}_j)^2)^{-1} \cdot d_{\mathbb{D}}(\mathbf{p}_i, \mathbf{p}_j) \nabla_i d_{\mathbb{D}}(\mathbf{p}_i, \mathbf{p}_j),$

**Descent Step:**  $R(\mathbf{p}_i - \eta \nabla_i \mathcal{L}) = \text{proj} \left( \mathbf{p}_i - \eta \frac{(1 - \|\mathbf{p}_i\|^2)^2}{4} \nabla_i \mathcal{L} \right),$

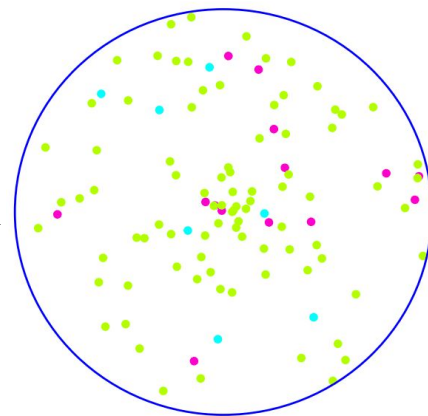
**Retraction Step:**  $\text{proj}(\mathbf{p}_i) = \begin{cases} \mathbf{p}_i / (\|\mathbf{p}_i\| + \varepsilon) & \text{if } \|\mathbf{p}_i\| \geq 1 \\ \mathbf{p}_i & \text{otherwise,} \end{cases}$

# GrassCaré

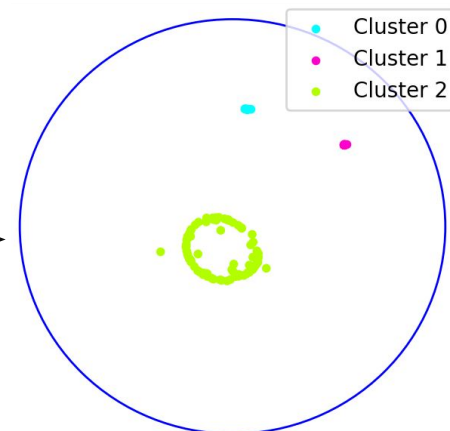
$$\min_{\mathbf{p}_1, \dots, \mathbf{p}_N} \text{KL}(\mathbf{P}_G || \mathbf{P}_D)$$



Input



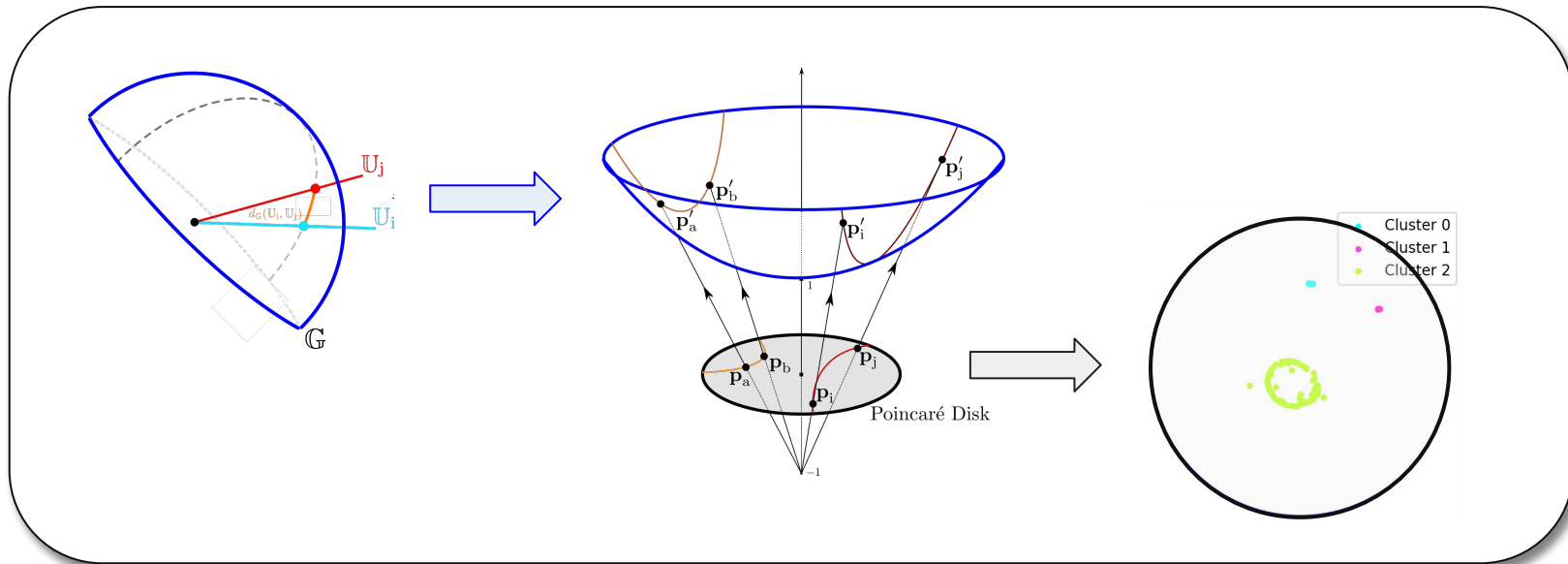
Optimizing



Output



# Question: Is GrassCaré any good?



# Theoretical Results: GrassCaré is accurate for clusters!

**Theorem 1.** Suppose  $N > 3$ . Define  $\gamma := \min_i \gamma_i$  and  $\Gamma := \max_i \gamma_i$ . Let  $\{\mathcal{U}_1, \dots, \mathcal{U}_K\}$  be a partition of  $\{\mathbb{U}_1, \dots, \mathbb{U}_N\}$  such that  $|\mathcal{U}_k| \geq n_K > 1 \forall k$ . Let

(Reverse) cluster density  $\delta := \frac{1}{\sqrt{2}\gamma} \max_k \max_{\mathbb{U}_i, \mathbb{U}_j \in \mathcal{U}_k} d_G(\mathbb{U}_i, \mathbb{U}_j),$  Distance between clusters  $\Delta := \frac{1}{\sqrt{2}\Gamma} \min_{\substack{\mathbb{U}_i \in \mathcal{U}_k, \mathbb{U}_j \in \mathcal{U}_\ell \\ k \neq \ell}} d_G(\mathbb{U}_i, \mathbb{U}_j).$

Then the optimal loss of GrassCaré is bounded by:

$$\mathcal{L}^* < \log D + \frac{5e^{\delta^2 - \Delta^2}}{(n_K - 1)},$$

Low for clusters!

where

$$D := \boxed{N}(n_K - 1) + N(N - \boxed{n_K}) \cdot \exp\left(-\operatorname{arcosh}^2\left(1 + \frac{2 \sin(\pi/\boxed{K})}{0.75^2}\right)\right). \quad (7)$$

Number of subspaces      Size of the smallest cluster      Number of clusters



# Experimental Results

# Experiment (3D)

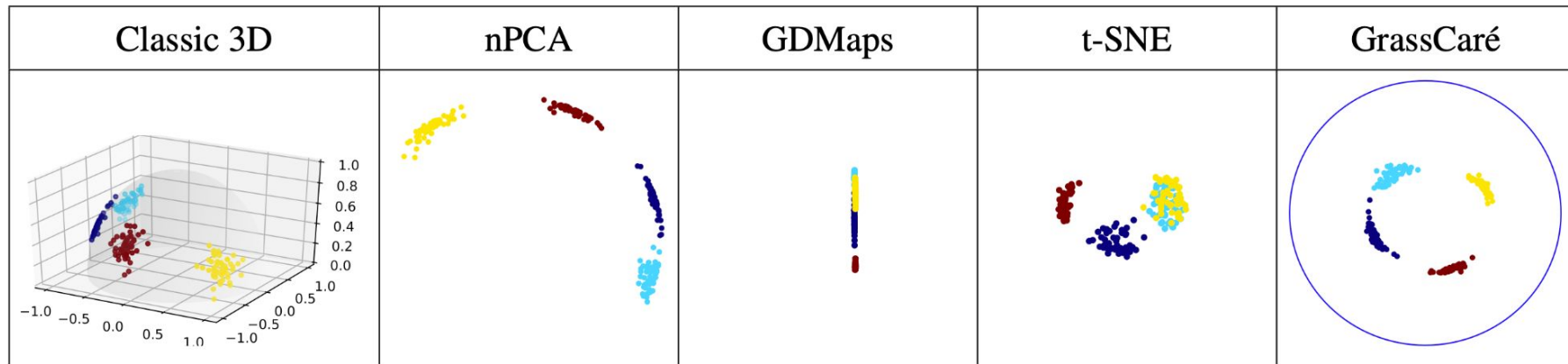


Figure 4: Visualizations of clusters in  $\mathbb{G}(3,1)$  with 4 clusters. GrassCaré produces a more accurate representation of the Grassmannian. nPCA and even the 3D representation display Clusters 1 and 2 (cyan and yellow) nearly diametrically apart. In reality they are quite close, as depicted by GrassCaré.

# Experiment (Representation Error)

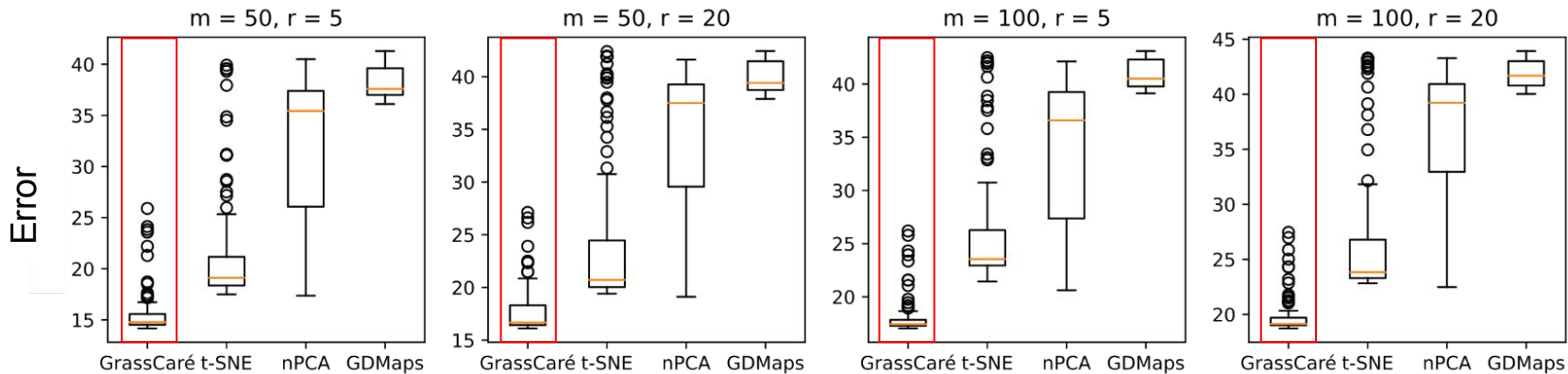


Figure 5: Representation error of GrassCaré (this paper) and other methods for high-dimensional Grassmannians.

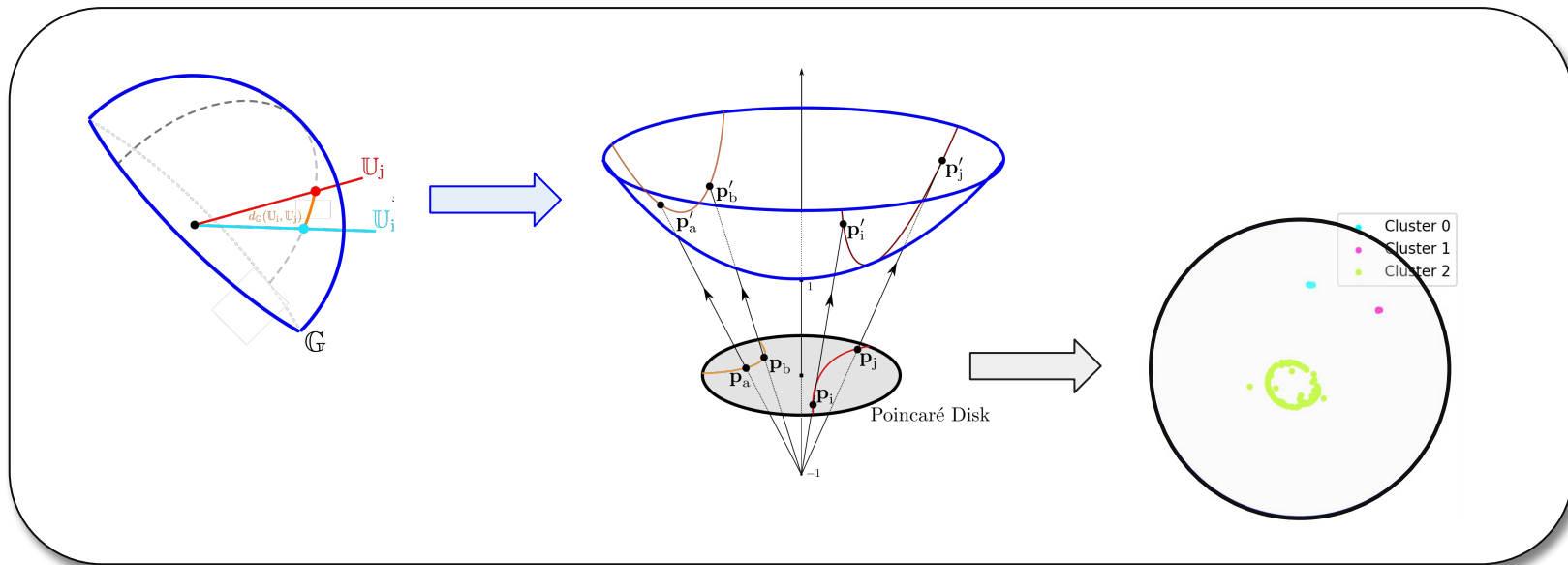
**Error on Poincaré Disk:**

$$\epsilon^2(\mathbb{D}) = \sum_{i,j} \left( \frac{d_{\mathbb{G}}(\mathbf{U}_i, \mathbf{U}_j)^2}{\sum d_{\mathbb{G}}(\mathbf{U}_i, \mathbf{U}_j)^2} - \frac{d_{\mathbb{D}}(\mathbf{p}_i, \mathbf{p}_j)^2}{\sum d_{\mathbb{D}}(\mathbf{p}_i, \mathbf{p}_j)^2} \right)$$

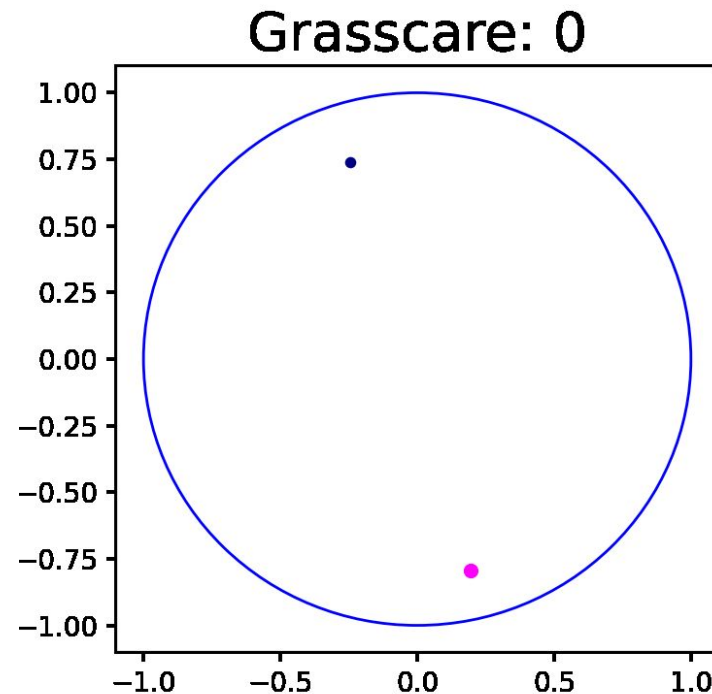
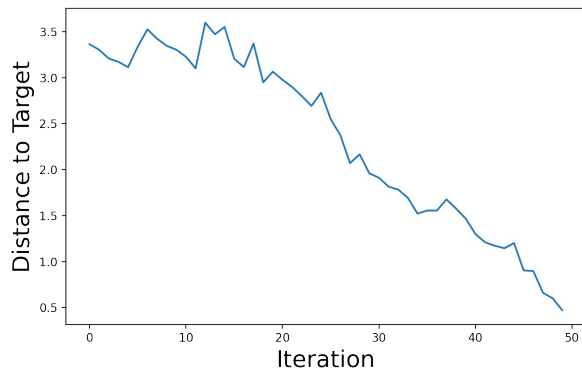
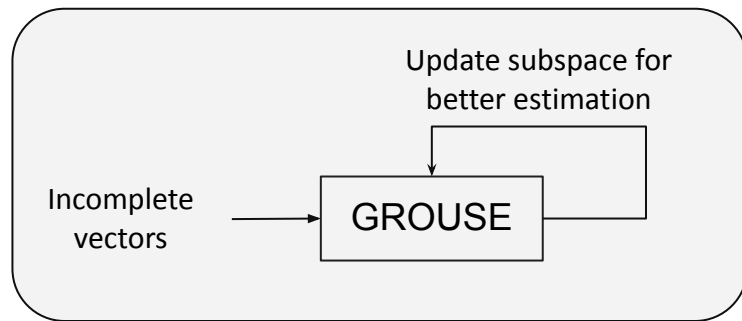
**Error on Euclidean:**

$$\epsilon^2(\mathbb{V}) = \sum_{i,j} \left( \frac{d_{\mathbb{G}}(\mathbf{U}_i, \mathbf{U}_j)^2}{\sum d_{\mathbb{G}}(\mathbf{U}_i, \mathbf{U}_j)^2} - \frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{\sum \|\mathbf{v}_i - \mathbf{v}_j\|^2} \right)$$

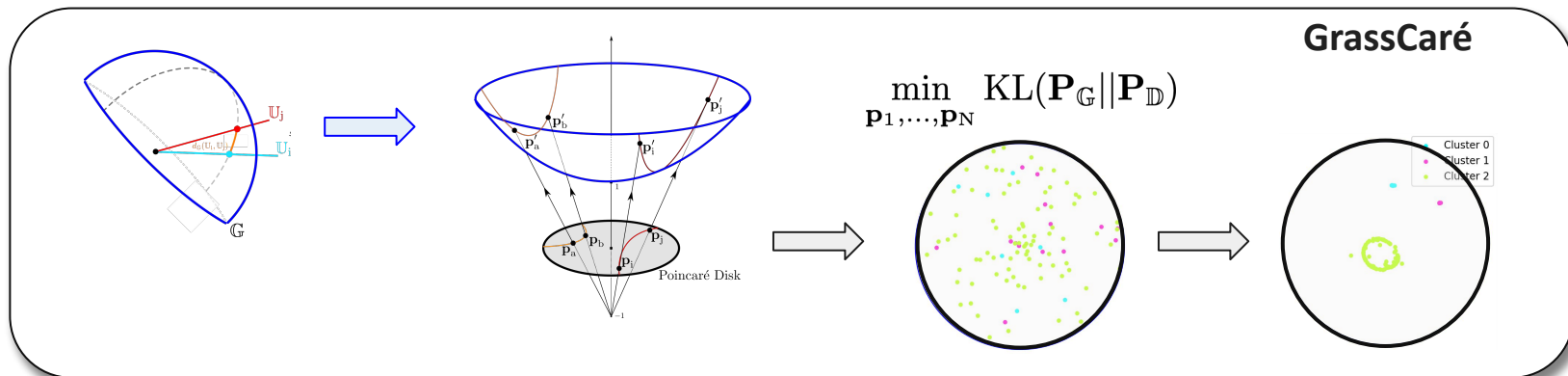
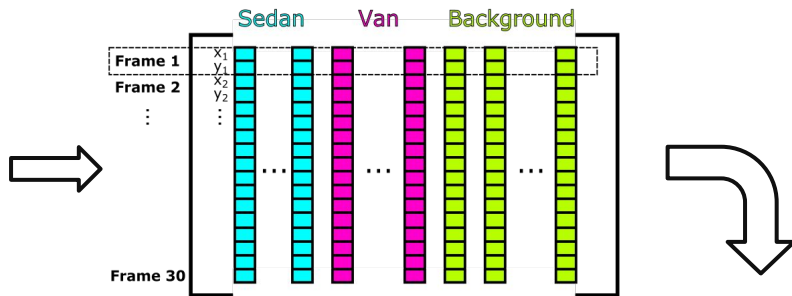
# Next question: What else can this be used for?



# Experiment (Subspace Estimation from Incomplete data)



# Conclusion







## Current Limitations / Future Steps

- Speed up the computation for larger dataset ( $>1k$  subspaces).
- Derive better theoretical bounds for non-cluster structures.

**Thanks!**

