GrassCaré

Visualizing **Grass**mannians via Poin**caré** Embeddings

- Huanran Li, Daniel Pimentel-Alarcón

What is a Grassmannian?

Grassmannian

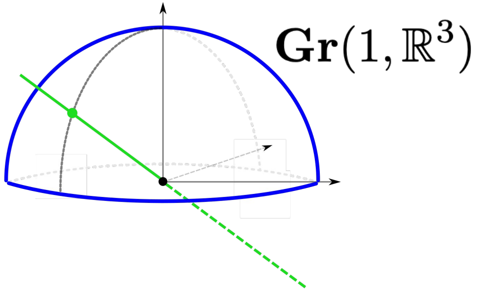
文A 12 languages

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.[1][2]

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



Recommender System

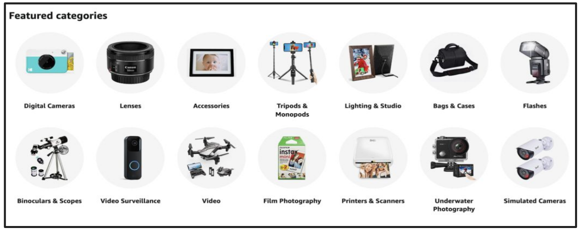
Type of Users => Subspace

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

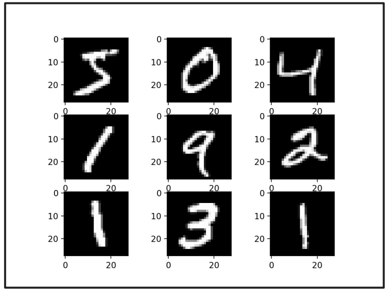
amazon



Object Classification

Class of Image => Subspace

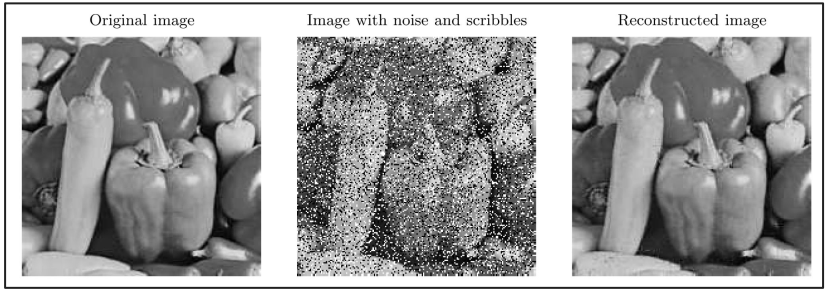
MNIST



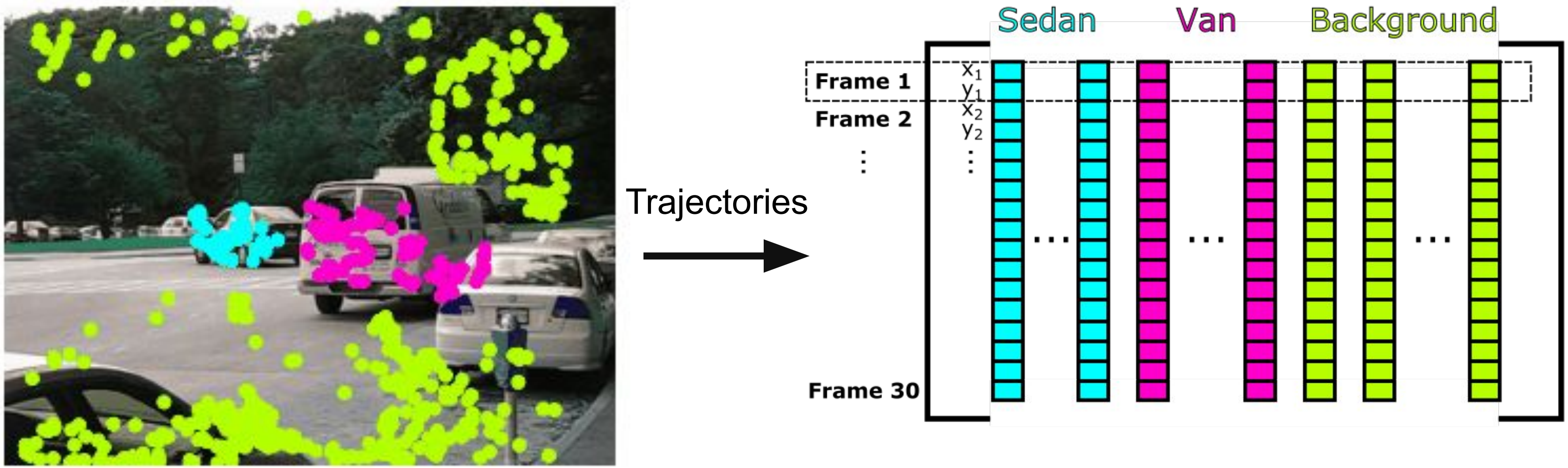
COIL-20

Image reconstruction

Color Scheme => Subspace

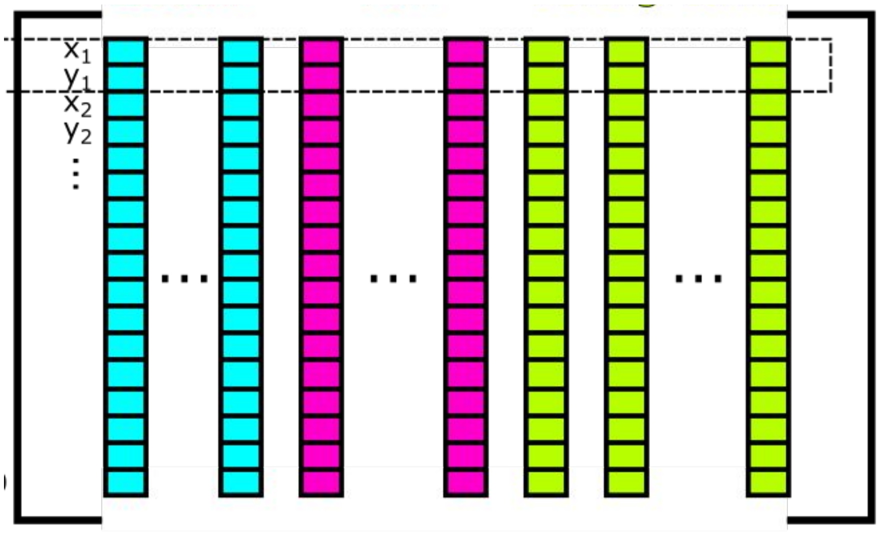


Object Trajectory



Sample's Pattern

Subspace 1 **Subspace 2** **Subspace 3**



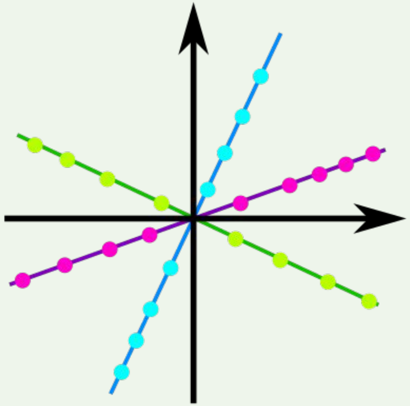
How **many** subspaces?

How **similar** they are?

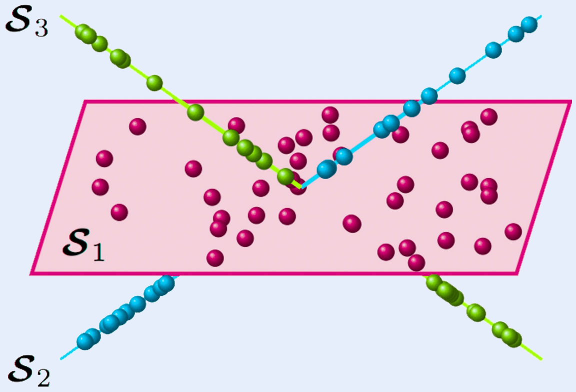
How **many samples** they have?

Motivation: Visualize Similarity between Subspace

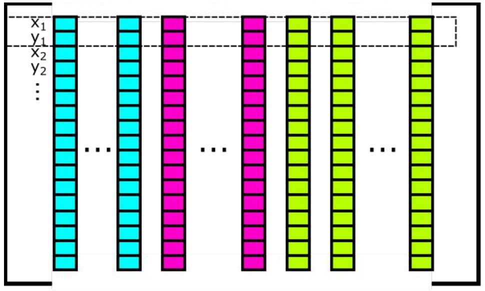
In 2-d Space



In 3-d Space

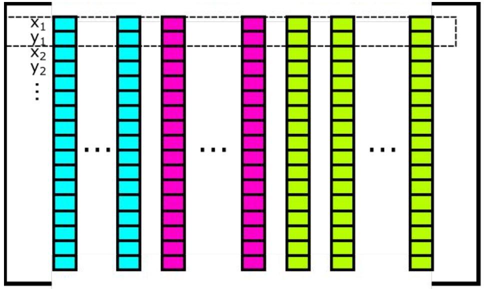


In 30-d Space



Elhamifar, Ehsan, and René Vidal. "Sparse subspace clustering: Algorithm, theory, and applications." *IEEE transactions on pattern analysis and machine intelligence* 35.11 (2013): 2765-2781.

How can we visualize High-dimensional subspaces?

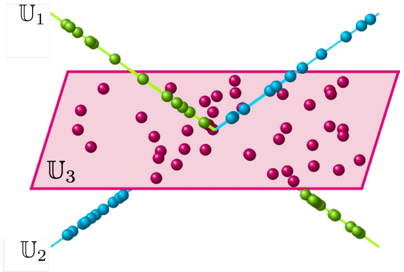


I SEE A ~~PATTERN~~ MATRIX, BUT MY IMAGINATION CANNOT PICTURE THE ~~MAKER~~ SUBSPACE OF THAT ~~PATTERN~~ MATRIX.

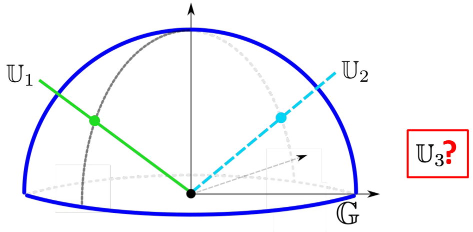
-- ALBERT EINSTEIN IN ~~1915~~ 2023

Existing options: Hemisphere Embedding

- **Given**: 1-d subspaces in 3-d space.



- **Output**: Point embedding on 3-d hemisphere.

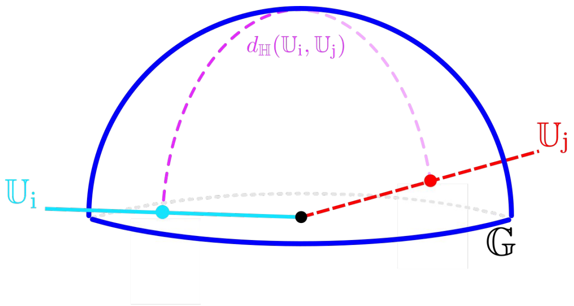


Drawback #1: limited to 1-d subspaces (aka lines).
Very restrictive!

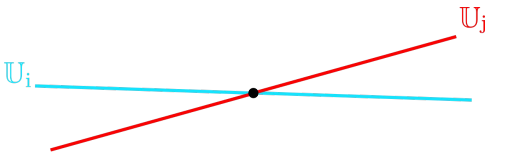
Existing options: Hemisphere Embedding

Drawback #2: Distance can be misleading!

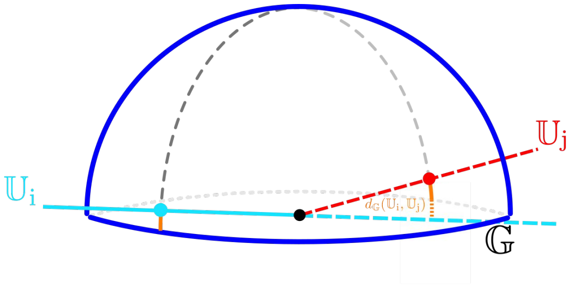
Subspaces Input



Visual Intuition



True distance



Distance between Subspaces:

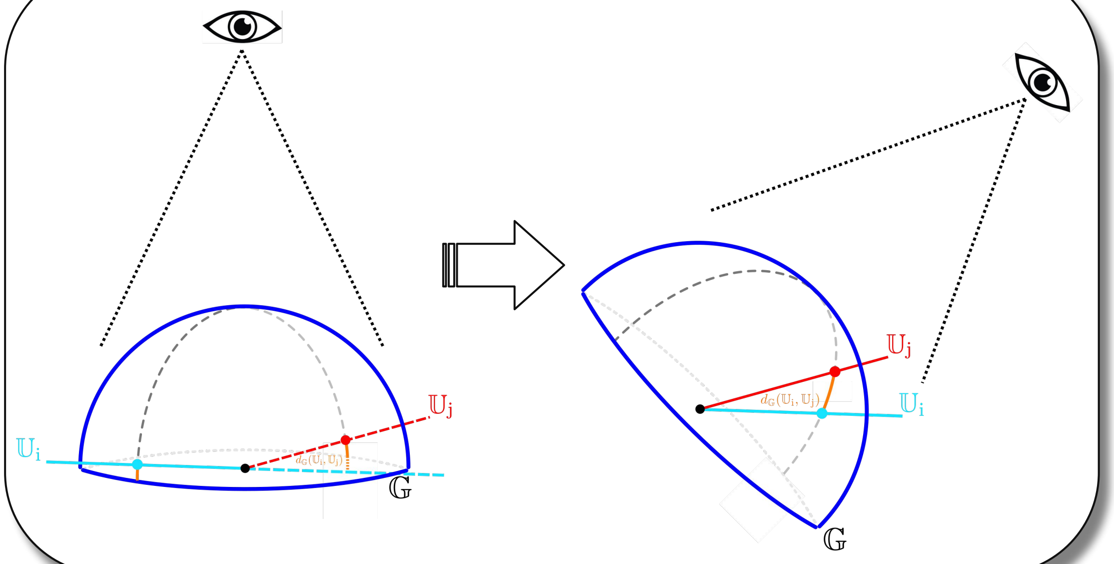
$$d_{\mathbb{G}}(\mathbb{U}_i, \mathbb{U}_j) := \sqrt{\sum_{\ell=1}^{r} \arccos^2 \sigma_\ell(\mathbf{U}_i^T \mathbf{U}_j)}$$

Basis of the subspaces

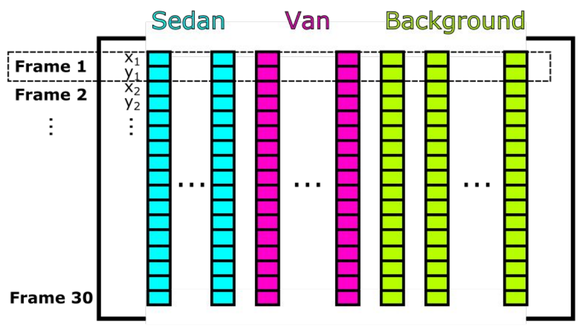
l'th largest singular value

2-Norm of the principle angle

Goal

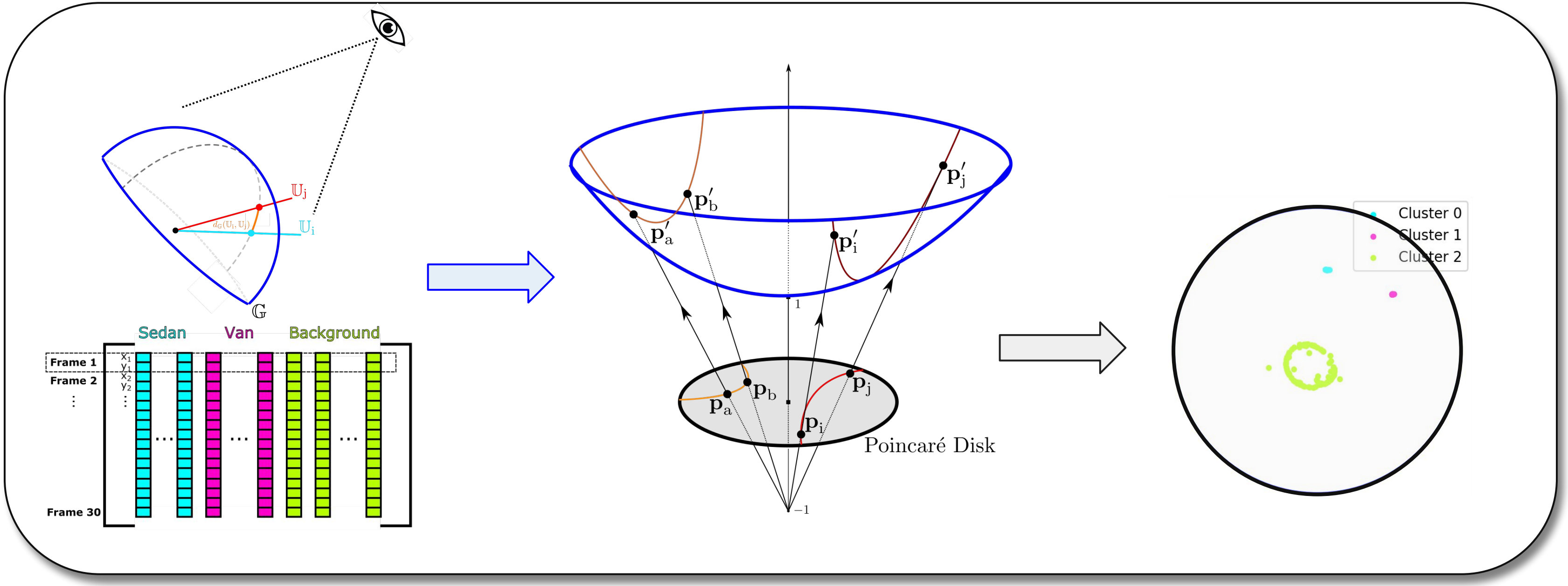


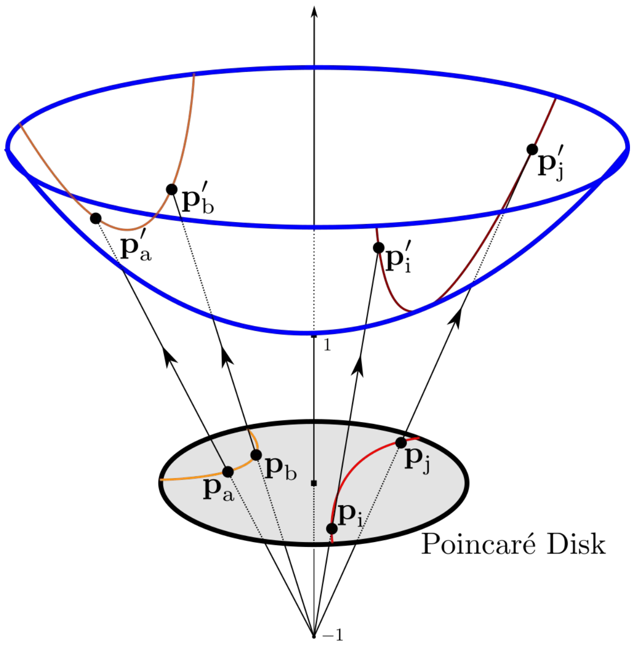
1)Find the right angle



2) high-dimensional subspaces

How do we find such embedding?





Distance between points in Poincaré Disk:

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

Euclidean Distance

Scaled by how far from origin

Distances near the edge are larger than they appear.

https://videohive.net/item/magnifying-glass/9013621

Problem Formulation

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

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

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

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

- Objective: The embedding preserves the most information.

$$\min_{\mathbf{p}_1, \ldots, \mathbf{p}_N} \mathrm{KL}(\boxed{\mathbf{P}_{\mathbb{G}}} \| \boxed{\mathbf{P}_{\mathbb{D}}})$$

Probability Matrix in Grassmannian   Probability Matrix in Poincaré disk

Probability that $\mathbb{U}_\mathrm{i}$ choose $\mathbb{U}_\mathrm{j}$ as the **closest neighbor.**

$$[\mathbf{P}_{\mathbb{G}}]_{\mathrm{ij}} := \frac{1}{2\mathrm{N}} \boxed{\frac{\exp(-\boxed{d_{\mathbb{G}}(\mathbb{U}_\mathrm{i}, \mathbb{U}_\mathrm{j})^2}/2\gamma_\mathrm{i}^2)}{\sum_{\mathrm{k}\neq\mathrm{i}} \exp(-d_{\mathbb{G}}(\mathbb{U}_\mathrm{i}, \mathbb{U}_\mathrm{k})^2/2\gamma_\mathrm{i}^2)}} + \boxed{\frac{1}{2\mathrm{N}} \frac{\exp(-d_{\mathbb{G}}(\mathbb{U}_\mathrm{j}, \mathbb{U}_\mathrm{i})^2/2\gamma_\mathrm{j}^2)}{\sum_{\mathrm{k}\neq\mathrm{j}} \exp(-d_{\mathbb{G}}(\mathbb{U}_\mathrm{j}, \mathbb{U}_\mathrm{k})^2/2\gamma_\mathrm{j}^2)}} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$$

Softmin function Subspace distance For symmetric purpose

$$\min_{\mathbf{p}_1,\ldots,\mathbf{p}_N} \mathrm{KL}(\mathbf{P}_{\mathbb{G}}||\mathbf{P}_{\mathbb{D}}) = \sum_{\mathrm{i,j}\in[1,\mathrm{N}]} [\mathbf{P}_{\mathbb{G}}]_{\mathrm{ij}} \log\left(\frac{[\mathbf{P}_{\mathbb{G}}]_{\mathrm{ij}}}{[\mathbf{P}_{\mathbb{D}}]_{\mathrm{ij}}}\right)$$

Softmin function Poincaré Disk distance

$$[\mathbf{P}_{\mathbb{D}}]_{\mathrm{ij}} := \boxed{\frac{\exp(-\boxed{d_{\mathbb{D}}(\mathbf{p}_\mathrm{i}, \mathbf{p}_\mathrm{j})^2}/\beta)}{\sum_{\mathrm{k}\neq l} \exp(-d_{\mathbb{D}}(\mathbf{p}_\mathrm{k}, \mathbf{p}_l)^2/\beta)}} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$$

Probability that $\mathbf{p}_\mathrm{i}$ choose $\mathbf{p}_\mathrm{j}$ as the **closest neighbor.**

Riemannian Gradient Descent

Gradient Descent Step: $\mathrm{p}_\mathrm{i}^{t+1} \leftarrow R(\mathrm{p}_\mathrm{i}^t - \eta \boldsymbol{\nabla}_\mathrm{i}\mathcal{L}),$

Gradient respect to Poincaré Distance: $\boldsymbol{\nabla}_\mathrm{i} d_\mathbb{D}(\mathrm{p}_\mathrm{i}, \mathrm{p}_\mathrm{j}) = \frac{4}{b\sqrt{c^2-1}}\left(\frac{||\mathrm{p}_\mathrm{j}||^2 - 2\langle \mathrm{p}_\mathrm{i}, \mathrm{p}_\mathrm{j}\rangle + 1}{a^2}\mathrm{p}_\mathrm{i} - \frac{\mathrm{p}_\mathrm{j}}{a}\right).$

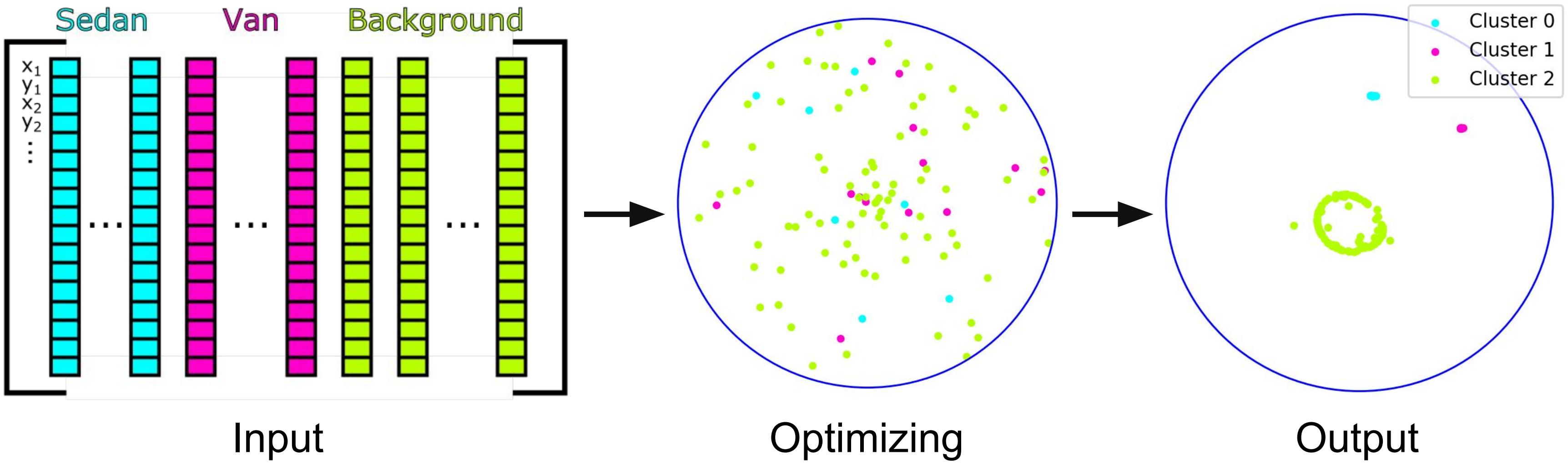
Gradient respect to Loss: $\boldsymbol{\nabla}_\mathrm{i}\mathcal{L} = \frac{4}{\beta}\sum_\mathrm{j}([\mathrm{P}_\mathbb{G}]_\mathrm{ij} - [\mathrm{P}_\mathbb{D}]_\mathrm{ij})(1 + d_\mathbb{D}(\mathrm{p}_\mathrm{i}, \mathrm{p}_\mathrm{j})^2)^{-1} \cdot d_\mathbb{D}(\mathrm{p}_\mathrm{i}, \mathrm{p}_\mathrm{j})\boldsymbol{\nabla}_\mathrm{i} d_\mathbb{D}(\mathrm{p}_\mathrm{i}, \mathrm{p}_\mathrm{j}),$

Descent Step: $R(\mathrm{p}_\mathrm{i} - \eta\boldsymbol{\nabla}_\mathrm{i}\mathcal{L}) = \mathrm{proj}\left(\mathrm{p}_\mathrm{i} - \eta\frac{(1-||\mathrm{p}_\mathrm{i}||^2)^2}{4}\boldsymbol{\nabla}_\mathrm{i}\mathcal{L}\right),$

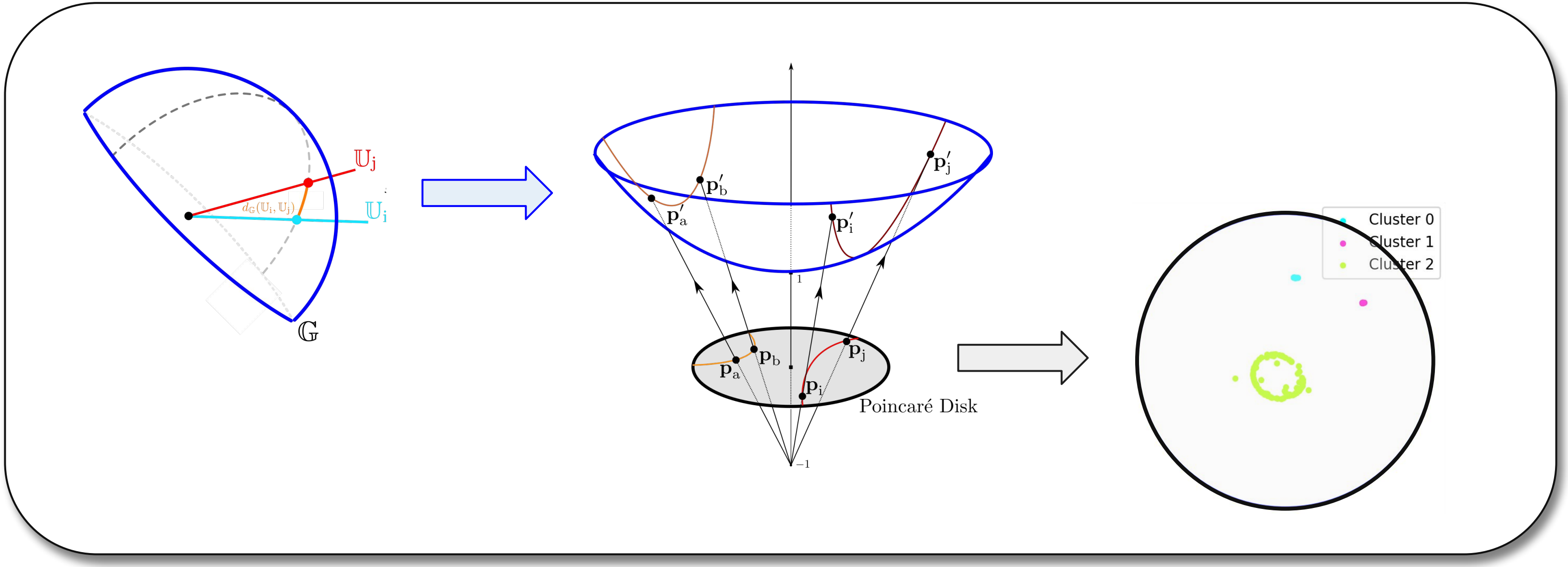
Retraction Step: $\mathrm{proj}(\mathrm{p}_\mathrm{i}) = \begin{cases} \mathrm{p}_\mathrm{i}/(||\mathrm{p}_\mathrm{i}|| + \varepsilon) & if \;\; ||\mathrm{p}_\mathrm{i}|| \geq 1 \\ \mathrm{p}_\mathrm{i} & \text{otherwise,} \end{cases}$

GrassCaré

$$\min_{\mathbf{p}_1,\dots,\mathbf{p}_N} \mathrm{KL}(\mathbf{P}_{\mathbb{G}}||\mathbf{P}_{\mathbb{D}})$$



Question:
Is GrassCaré any good?



Theoretical Results: GrassCaré is accurate for clusters!

Theorem 1. Suppose $\mathrm{N} > 3$. Define $\gamma := \min_{\mathrm{i}} \gamma_{\mathrm{i}}$ and $\Gamma := \max_{\mathrm{i}} \gamma_{\mathrm{i}}$. Let $\{\mathcal{U}_1, \ldots, \mathcal{U}_{\mathrm{K}}\}$ be a partition of $\{\mathbb{U}_1, \ldots, \mathbb{U}_{\mathrm{N}}\}$ such that $|\mathcal{U}_{\mathrm{k}}| \geq \mathrm{n_K} > 1 \; \forall \, \mathrm{k}$. Let

(Reverse) cluster **density**

$$\boxed{\delta := \frac{1}{\sqrt{2\gamma}} \max_{\mathrm{k}} \max_{\mathbb{U}_{\mathrm{i}}, \mathbb{U}_{\mathrm{j}} \in \mathcal{U}_{\mathrm{k}}} d_{\mathbb{G}}(\mathbb{U}_{\mathrm{i}}, \mathbb{U}_{\mathrm{j}}),} \quad \boxed{\Delta := \frac{1}{\sqrt{2\Gamma}} \min_{\substack{\mathbb{U}_{\mathrm{i}} \in \mathcal{U}_{\mathrm{k}}, \mathbb{U}_{\mathrm{j}} \in \mathcal{U}_{\ell}: \\ \mathrm{k} \neq \ell}} d_{\mathbb{G}}(\mathbb{U}_{\mathrm{i}}, \mathbb{U}_{\mathrm{j}}).}$$

Distance between clusters

Then the optimal loss of GrassCaré is bounded by:

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

Low for clusters!

where

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

Number of subspaces Size of the smallest cluster Number of clusters

Experimental Results

Experiment (3D)

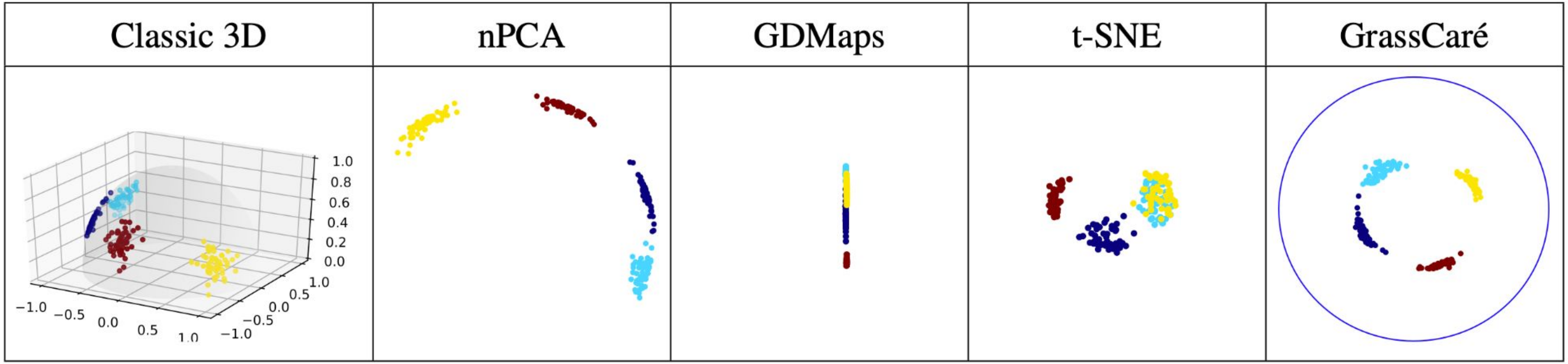


Figure 4: Visualizations of clusters in $\mathbb{G}(3,1)$ with 4 clusters. GrassCaré produces a more accurate representation of the Grassmannian. nPCA and even the 3D representation display Clusters 1 and 2 (cyan and yellow) nearly diametrically apart. In reality they are quite close, as depicted by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

Experiment (Representation Error)

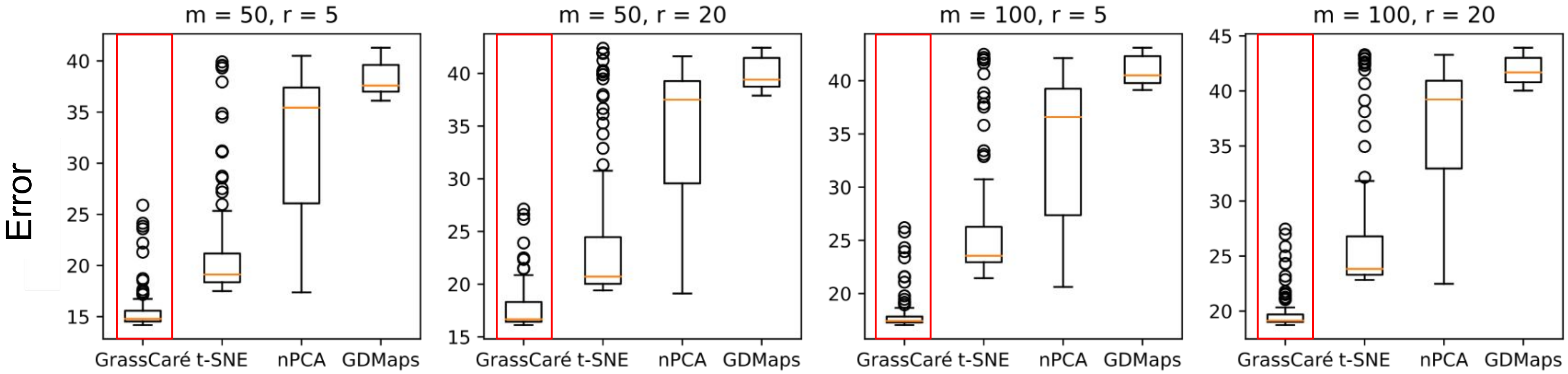
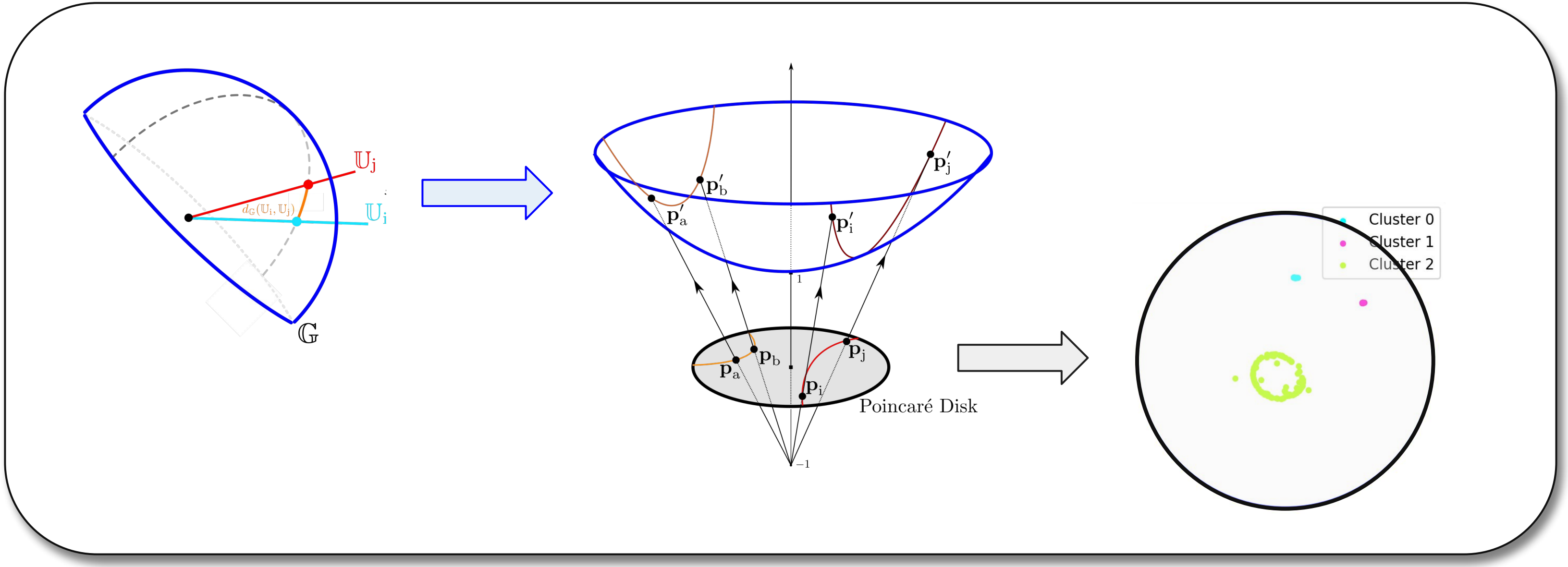


Figure 5: Representation error of GrassCaré (this paper) and other methods for high-dimensional Grassmannians.

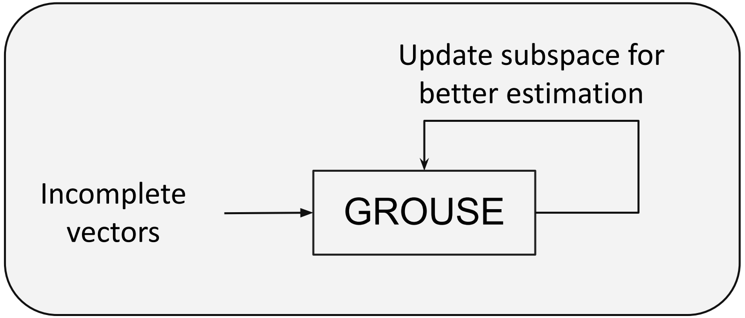
Error on Poincaré Disk: $$\epsilon^2(\mathbb{D}) = \sum_{i,j}\left(\frac{d_{\mathbb{G}}(\mathbb{U}_i, \mathbb{U}_j)^2}{\sum d_{\mathbb{G}}(\mathbb{U}_i, \mathbb{U}_j)^2} - \frac{d_{\mathbb{D}}(\mathbf{p}_i, \mathbf{p}_j)^2}{\sum d_{\mathbb{D}}(\mathbf{p}_i, \mathbf{p}_j)^2}\right)$$

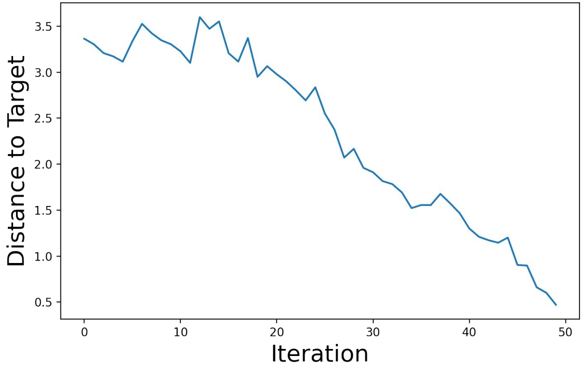
Error on Euclidean: $$\epsilon^2(\mathbb{V}) = \sum_{i,j}\left(\frac{d_{\mathbb{G}}(\mathbb{U}_i, \mathbb{U}_j)^2}{\sum d_{\mathbb{G}}(\mathbb{U}_i, \mathbb{U}_j)^2} - \frac{||\mathbf{v}_i - \mathbf{v}_j||^2}{\sum ||\mathbf{v}_i - \mathbf{v}_j||^2}\right)$$

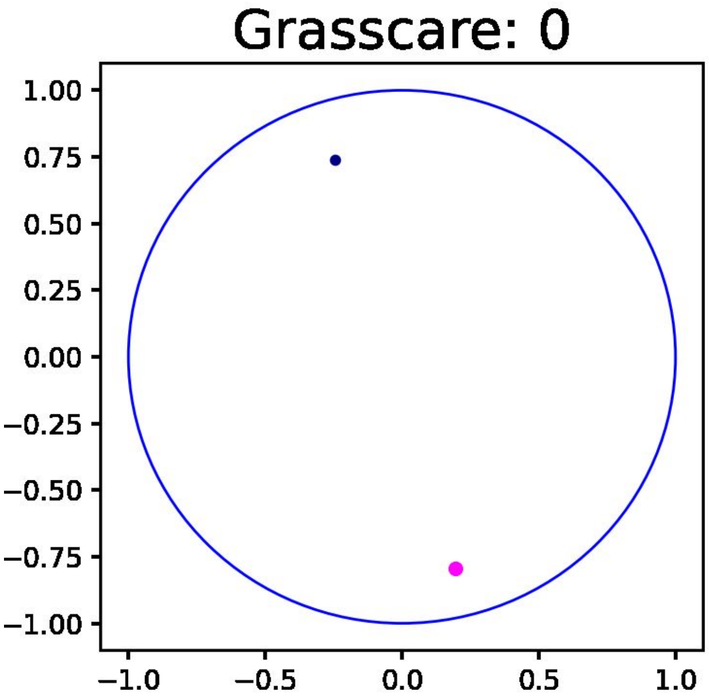
Next question:
What else can this be used for?



Experiment (Subspace Estimation from Incomplete data)

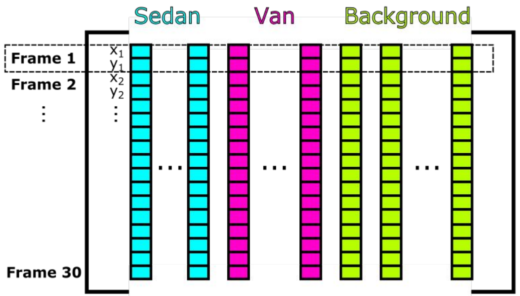


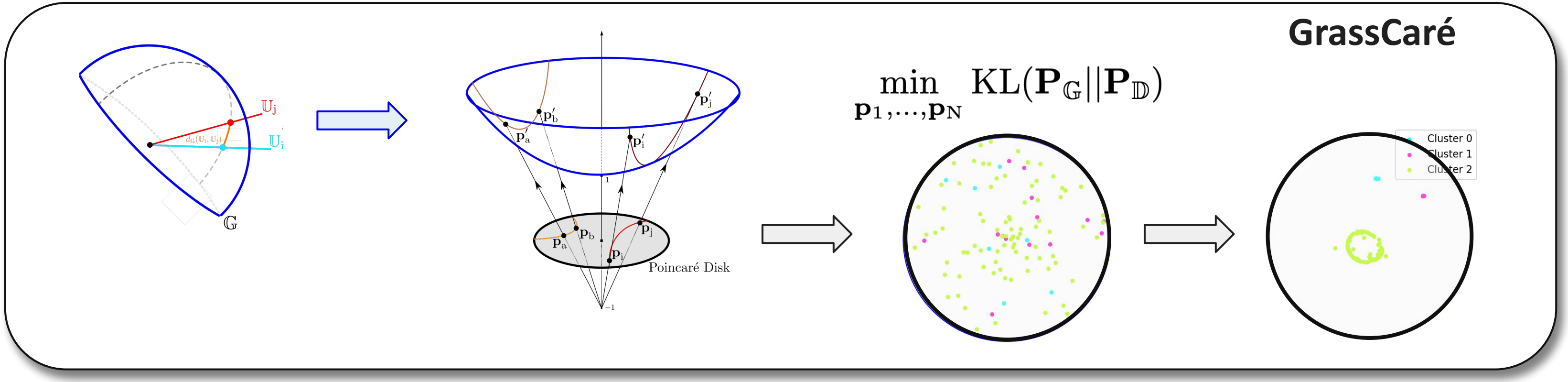




Balzano, L., Nowak, R., and Recht, B. Online identification and tracking of subspaces from highly incomplete information. In *2010 48th Annual allerton conference on communication, control, and computing (Allerton)*, pp. 704–711. IEEE, 2010

Conclusion


Sedan
Van
Background
Frame 1
Frame 2
Frame 30


GrassCaré
$\min_{\mathbf{p}_1,\ldots,\mathbf{p}_N} \mathrm{KL}(\mathbf{P}_{\mathbb{G}}||\mathbf{P}_{\mathbb{D}})$
Poincaré Disk
Cluster 0
Cluster 1
Cluster 2

Current Limitations / Future Steps

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

Thanks!