Manifold clustering in non-Euclidean spaces

Xu Wang [1]   Konstantinos Slavakis [2]   Gilad Lerman [1]

[1]Department of Mathematics, University of Minnesota

[2]Department of ECE and Digital Technology Center, University of Minnesota

February 4, 2015

Motivation

Examples	Non-Euclidean data representation
Image texture	Symmetric positive definite matrix
Linear dynamic system	Grassmannian (subspaces)
Shape of 2D (3D) object	Shape space
$\cdots$	Stiefel, $SE(3)$, Lie group etc.

Motivation

Examples	Non-Euclidean data representation
Image texture	Symmetric positive definite matrix
Linear dynamic system	Grassmannian (subspaces)
Shape of 2D (3D) object	Shape space
$\cdots$	Stiefel, $SE(3)$, Lie group etc.

- **Goal**: Cluster such data sets (especially when clusters lie on low-dimensional submanifolds that may intersect)

Clustering for Euclidean vectors

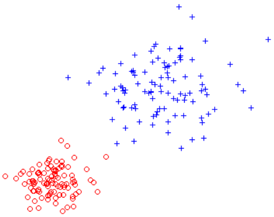


Figure : K-means

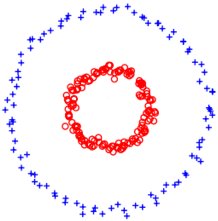
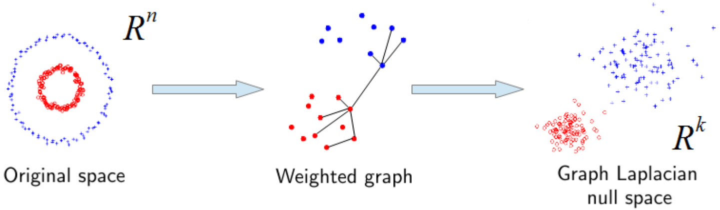


Figure : Spectral clustering

Spectral clustering

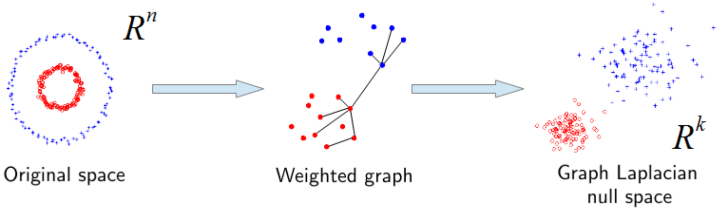
Spectral clustering contains two steps:



- weights $A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2}$

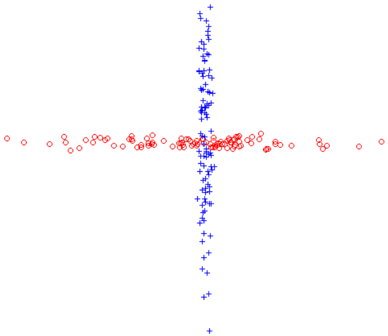
Spectral clustering

Spectral clustering contains two steps:

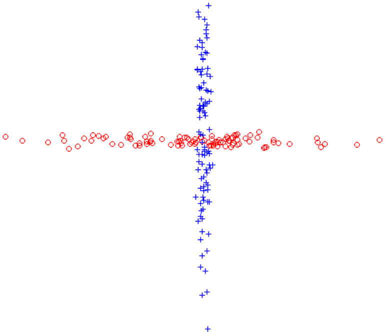


- weights $A_{ij} = e^{-d^2(x_i, x_j)/\sigma^2}$
- $d(x_i, x_j)$ can be any metric. This leads to a version of **spectral clustering with Riemannian metric** (SCR)

Hybrid linear modeling (subspace clustering)

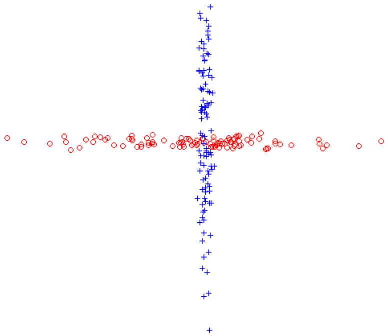


Hybrid linear modeling (subspace clustering)



- weights $A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2}$

Hybrid linear modeling (subspace clustering)



- weights $A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2}$
- Methods (e.g., SCR) with only distance information, fail at the intersection!

A clustering algorithm: Sparse subspace clustering

- For each point $\mathbf{x}_i$, solve the following sparse optimization

$$\min \sum_{j \neq i} |w_{ij}| + \lambda \|\mathbf{x}_i - \sum_{j \neq i} w_{ij} \mathbf{x}_j\|^2 \qquad s.t. \sum_{j \neq i} w_{ij} = 1$$

A clustering algorithm: Sparse subspace clustering

- For each point $\mathbf{x}_i$, solve the following sparse optimization

$$\min \sum_{j \neq i} |w_{ij}| + \lambda \|\mathbf{x}_i - \sum_{j \neq i} w_{ij}\mathbf{x}_j\|^2 \qquad s.t. \sum_{j \neq i} w_{ij} = 1$$

- The top nonzero coefficients come from points in the same subspace

A clustering algorithm: Sparse subspace clustering

- For each point $\mathbf{x}_i$, solve the following sparse optimization

$$\min \sum_{j \neq i} |w_{ij}| + \lambda \|\mathbf{x}_i - \sum_{j \neq i} w_{ij} \mathbf{x}_j\|^2 \qquad s.t. \sum_{j \neq i} w_{ij} = 1$$

- The top nonzero coefficients come from points in the same subspace
- $A_{ij} = |w_{ij}| + |w_{ji}|$

Naive generalization: Sparse manifold clustering (SMC)

For each point $\mathbf{x}_i$, solve the following sparse optimization

$$\min \sum_{j\neq i} |w_{ij}| + \lambda \|\log_{\mathbf{x}_i} \mathbf{x}_i - \sum_{j\neq i} w_{ij} \log_{\mathbf{x}_i} \mathbf{x}_j\|^2$$

- Linearization: logarithm map $\log_{\mathbf{x}_i}$ maps all points to the tangent space $T_{\mathbf{x}_i}$ at $\mathbf{x}_i$

Naive generalization: Sparse manifold clustering (SMC)

For each point $\mathbf{x}_i$, solve the following sparse optimization

$$\min \sum_{j \neq i} |w_{ij}| + \lambda \|\log_{\mathbf{x}_i} \mathbf{x}_i - \sum_{j \neq i} w_{ij} \log_{\mathbf{x}_i} \mathbf{x}_j\|^2$$

- Linearization: logarithm map $\log_{\mathbf{x}_i}$ maps all points to the tangent space $T_{\mathbf{x}_i}$ at $\mathbf{x}_i$
- Limitation: this linearization introduces a lot of error when $\mathbf{x}_i$ and $\mathbf{x}_j$ are far away.

Naive generalization: Sparse manifold clustering (SMC)

For each point $\mathbf{x}_i$, solve the following sparse optimization

$$\min \sum_{j \neq i} |w_{ij}| + \lambda \| \log_{\mathbf{x}_i} \mathbf{x}_i - \sum_{j \neq i} w_{ij} \log_{\mathbf{x}_i} \mathbf{x}_j \|^2$$

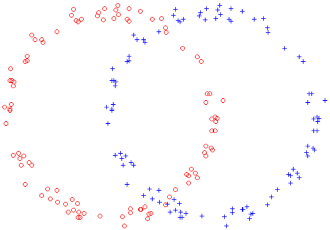
- Linearization: logarithm map $\log_{\mathbf{x}_i}$ maps all points to the tangent space $T_{\mathbf{x}_i}$ at $\mathbf{x}_i$
- Limitation: this linearization introduces a lot of error when $\mathbf{x}_i$ and $\mathbf{x}_j$ are far away.
- No guarantee! The top nonzero coefficients may not come from points in the same cluster.

back

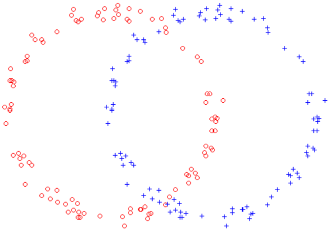
Manifold clustering algorithms

- SCR: $A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2}$ (trouble at the intersection!)
- SMC: $A_{ij} = |w_{ij}| + |w_{ji}|$ (no guarantee for manifolds!)
- GCT (resolving intersection, theoretical guarantee)
- GCT stands for Geodesic Clustering with Tangents

The local PCA algorithm



The local PCA algorithm



- Multi-manifold model
- $A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2} e^{-\|\mathbf{C}_i - \mathbf{C}_j\|^2/\eta^2}$ where $\mathbf{C}_i$ is the covariance matrix computed from points in a neighborhood of x_i.

Generalization of local PCA to Riemannian spaces

How to generalize it to Riemannian manifolds?

$$A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2} e^{\|\mathbf{C}_i - \mathbf{C}_j\|^2/\eta^2}$$

Generalization of local PCA to Riemannian spaces

How to generalize it to Riemannian manifolds?

$$A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2} e^{\|\mathbf{C}_i - \mathbf{C}_j\|^2/\eta^2}$$

- what is the covariance matrix of a set $\{x_1, ..., x_n\}$ in Riemannian spaces?

Generalization of local PCA to Riemannian spaces

How to generalize it to Riemannian manifolds?

$$A_{ij} = e^{-d^2(x_i, x_j)/\sigma^2} e^{\|\mathbf{C}_i - \mathbf{C}_j\|^2/\eta^2}$$

- what is the covariance matrix of a set $\{x_1, ..., x_n\}$ in Riemannian spaces?
  - $\mathbf{C}_{x_i}$: Covariance of the vectors $\log_{x_i} x_1, ..., \log_{x_i} x_n$ in the tangent space T_{x_i}

Generalization of local PCA to Riemannian spaces

How to generalize it to Riemannian manifolds?

$$A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2} e^{\|\mathbf{C}_i - \mathbf{C}_j\|^2/\eta^2}$$

- what is the covariance matrix of a set $\{x_1, ..., x_n\}$ in Riemannian spaces?
  - $\mathbf{C}_{x_i}$: Covariance of the vectors $\log_{x_i} x_1, ..., \log_{x_i} x_n$ in the tangent space T_{x_i}
- How to compute the difference of $\mathbf{C}_i$ and $\mathbf{C}_j$?

Generalization of local PCA to Riemannian spaces

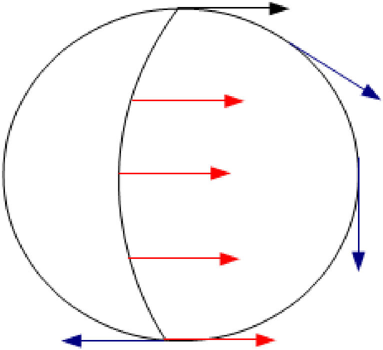
How to generalize it to Riemannian manifolds?

$$A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2} e^{\|\mathbf{C}_i - \mathbf{C}_j\|^2/\eta^2}$$

- what is the covariance matrix of a set $\{x_1, \ldots, x_n\}$ in Riemannian spaces?
  - $\mathbf{C}_{x_i}$: Covariance of the vectors $\log_{x_i} x_1, \ldots, \log_{x_i} x_n$ in the tangent space T_{x_i}
- How to compute the difference of $\mathbf{C}_i$ and $\mathbf{C}_j$?
  - (Caution!) $\mathbf{C}_i$ and $\mathbf{C}_j$ are quantities in different tangent spaces T_{x_i} and T_{x_j} and their values depend on the particular coordinate system chosen in each tangent space.

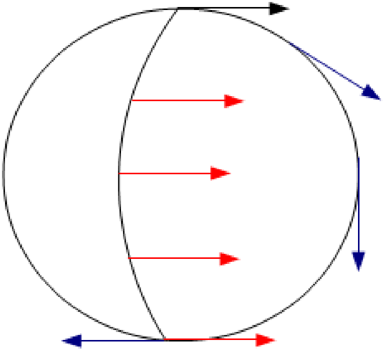
Generalization of local PCA to Riemannian spaces

Problem: identify vectors at the north and south poles



Generalization of local PCA to Riemannian spaces

Problem: identify vectors at the north and south poles



- Implication: can't compare $\mathbf{C}_i$ and $\mathbf{C}_j$ in a consistent way on $\mathbb{S}^2$!

Generalization of local PCA to Riemannian spaces

Theorem (Hairy ball theorem)

There is no nonvanishing continuous tangent vector field on any even-dimensional n-spheres, particularly, on $\mathbb{S}^2$.

Generalization of local PCA to Riemannian spaces

Theorem (Hairy ball theorem)

There is no nonvanishing continuous tangent vector field on any even-dimensional n-spheres, particularly, on $\mathbb{S}^2$.

Think about the hair whorl!

Theorem (Hairy ball theorem)

There is no nonvanishing continuous tangent vector field on any even-dimensional n-spheres, particularly, on $\mathbb{S}^2$.

Think about the hair whorl!

This is a special case of Poincaré-Hopf index theorem for general manifolds in differential topology. There is no hope to find nonzero vector fields on general manifolds, let alone consistent coordinate systems.

Generalization of local PCA to Riemannian spaces

How to generalize it to Riemannian manifolds?

$$A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2} e^{\|\mathbf{C}_i - \mathbf{C}_j\|^2/\eta^2}$$

Dead end?

Generalization of local PCA to Riemannian spaces

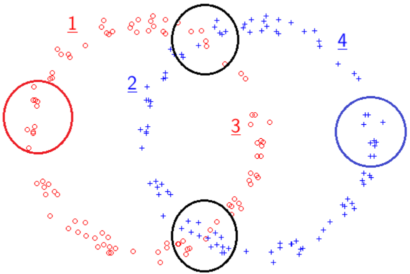
How to generalize it to Riemannian manifolds?

$$A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2} e^{\|\mathbf{C}_i - \mathbf{C}_j\|^2/\eta^2}$$

Dead end?

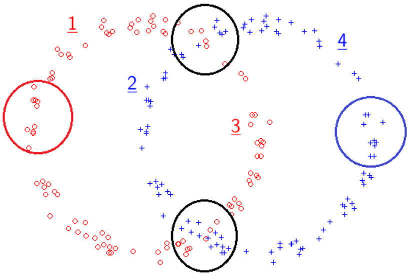
Problem: $\mathbf{C}_i$ depends on coordinate systems, in other words, "not intrinsic".

Solution: find coordinate-independent quantities! SMC



- Find the local dimension of the data by thresholding the top eigenvalues of the covariance matrix under any coordinate system.

$$A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2}\mathbf{1}_{\dim(x_i)=\dim(x_j)}$$



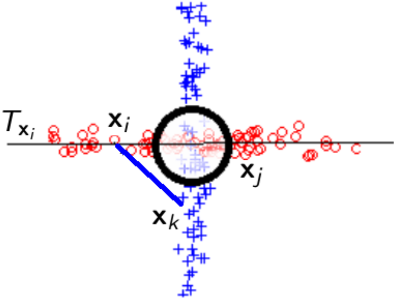
- Find the local dimension of the data by thresholding the top eigenvalues of the covariance matrix under any coordinate system.

$$A_{ij} = e^{-d^2(x_i,x_j)/\sigma^2}\mathbf{1}_{\dim(x_i)=\dim(x_j)}$$

Caution!

Geodesic Clustering with Tangents (GCT)

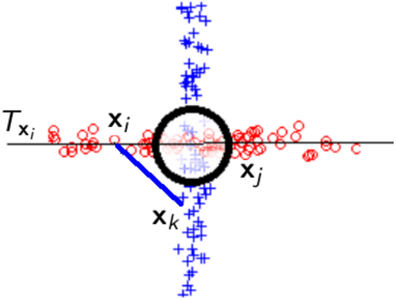
Quantities independent of coordinate systems



- $T_{\mathbf{x}_i}$ is the tangent space at point $\mathbf{x}_i$
- Find the geodesic angle θ_{ij} of any two points $\mathbf{x}_i$ and $\mathbf{x}_j$

Geodesic Clustering with Tangents (GCT)

Quantities independent of coordinate systems



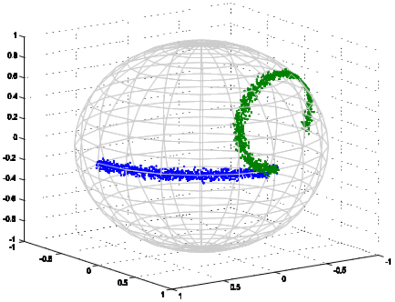
- $T_{\mathbf{x}_i}$ is the tangent space at point $\mathbf{x}_i$
- Find the geodesic angle θ_{ij} of any two points $\mathbf{x}_i$ and $\mathbf{x}_j$
- $\theta_{ij} \ll \theta_{ik}$

$$A_{ij} = e^{-d^2(\mathbf{x}_i,\mathbf{x}_j)/\sigma^2} \mathbf{1}_{\dim(\mathbf{x}_i)=\dim(\mathbf{x}_j)} e^{-(\theta_{ij}+\theta_{ji})/\eta} \gg A_{ik}$$

Theorem of GCT: assume the data points lie on two (geodesic) submanifolds of a general Riemannian manifold. With high probability and the proper choices of parameters specified in the paper, the constructed graph has two distinct major components and a few isolated nodes, where **each component** corresponds to **a cluster** of the original data points.

Experiment

Testing on synthetic dataset: Arc and spiral on $\mathbb{S}^2$



Methods	Clustering accuracy rate
GCT (proposed)	0.96
SMC	0.69
SCR	0.53

More tests on different manifolds can be found in the paper.

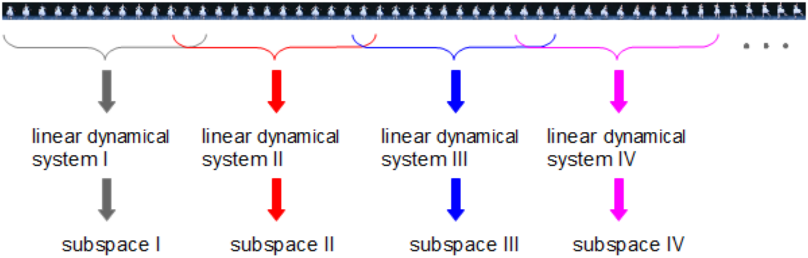
Experiment

Ballet dataset contains videos from a ballet instruction DVD.

Figure : Two samples of Ballet video sequences: The first and second rows comprise samples from the actions of hopping and leg-swinging, respectively.

Experiment

For a video, we generate a sequence of subspaces as follows.

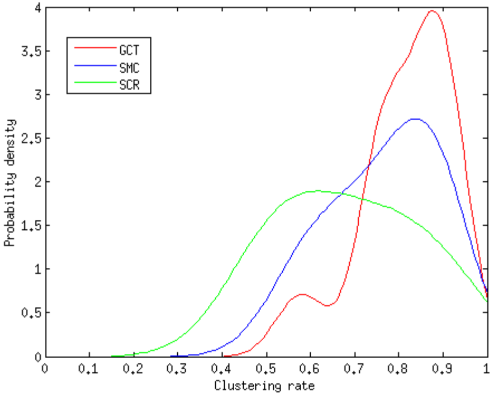


Experiment

For one dataset, we generate 3 clusters of subspaces from 3 random ballet videos. We do the experiment over 30 such datasets.

Experiment

For one dataset, we generate 3 clusters of subspaces from 3 random ballet videos. We do the experiment over 30 such datasets.



Summary

- We analyzed possible ways to cluster manifold data (e.g., SCR, SMC, GCT).
- SCR works well in general, but is not able to resolve intersections.
- SMC formally generalizes the SSC algorithm, but there is no theoretical guarantee.
- GCT (proposed) is theoretical guaranteed under multi-manifold model and able to deal with intersections.