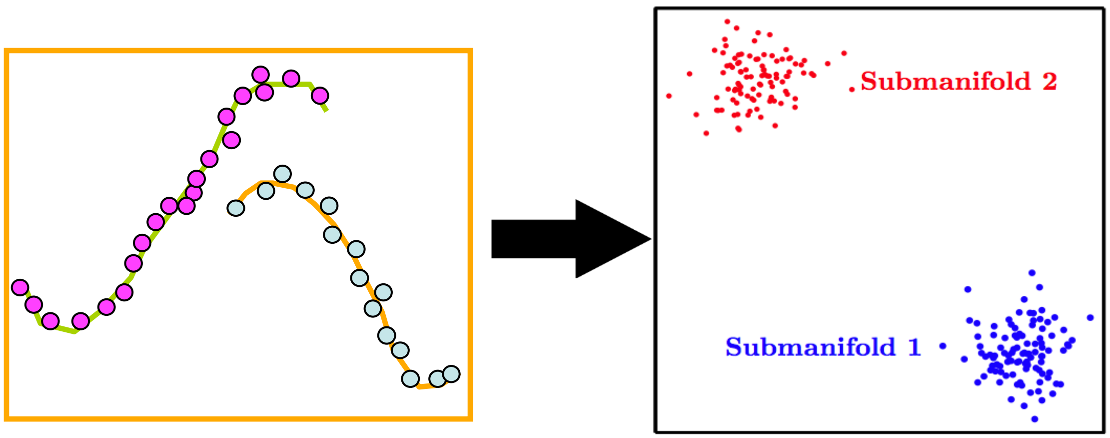
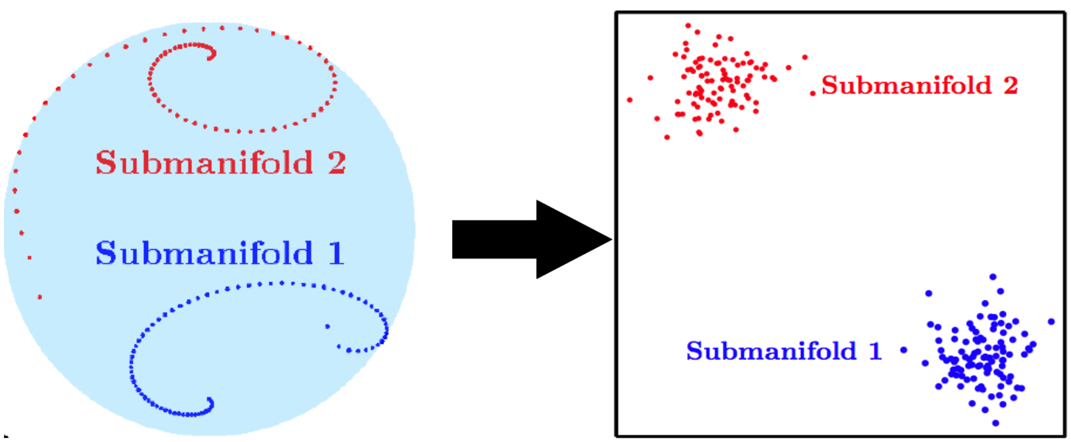
Locally Linear Manifold Clustering (LLMC)

- Nonlinear manifolds in a Euclidean space



- Nonlinear sub-manifolds in a Riemannian space

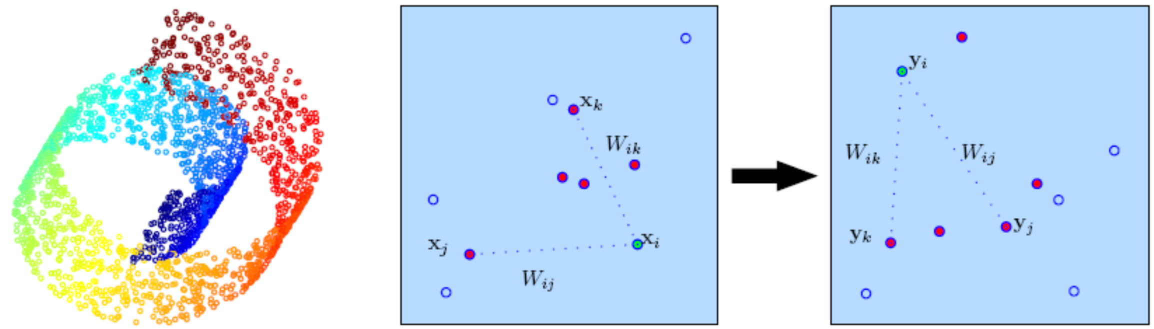


- Goals
  - Develop framework for simultaneous clustering & dimensionality reduction
  - Reduce manifold clustering to central clustering
- Contributions
  - Extend NLDR from one sub-manifold to multiple sub-manifolds
  - Extend NLDR from Euclidean spaces to Riemannian spaces
  - Show that when sub-manifolds are separated, all points in one sub-manifold are mapped to single point

Nonlinear dimension reduction & clustering

- Global techniques
  - Isomap (Tenenbaum et al. ‘00)
  - Kernel PCA (Schölkopf-Smola’98)
- Local techniques
  - Locally Linear Embedding (LLE) (Roweis-Saul ’00)
  - Laplacian Eigenmaps (LE) (Belkin-Niyogi ‘02)
  - Hessian LLE (HLLE) (Donoho-Grimes ‘03)
  - Local Tangent Space Alignment (Zha-Zhang’05)
  - Maximum Variance Unfolding (Weinberger-Saul ‘04)
  - Conformal Eigenmaps (Sha-Saul’05)
- Clustering based on geometry
  - LLE+Spectral clustering (Polito-Perona ’02)
  - Spectral embedding and clustering (Brand-Huang’03)
  - Isomap+EM (Souvenir-Pless’05)
- Clustering based on dimension
  - Fractal dimension (Barbara-Chen’00)
  - Tensor voting (Mordohai-Medioni’05)
  - Dimension induced clustering (Gionis et al. ’05)
  - Translated Poisson mixtures (Haro et al.’08)

Locally linear embedding (LLE)



(a) Original manifold (b) Learning matrix of weights (c) Low-dimensional embedding

- Find the k-nearest neighbors of each data point according to the Euclidean distance.

- Compute a matrix W that represents the local neighborhood as the affine subspace spanned by a point and its k-nearest neighbors

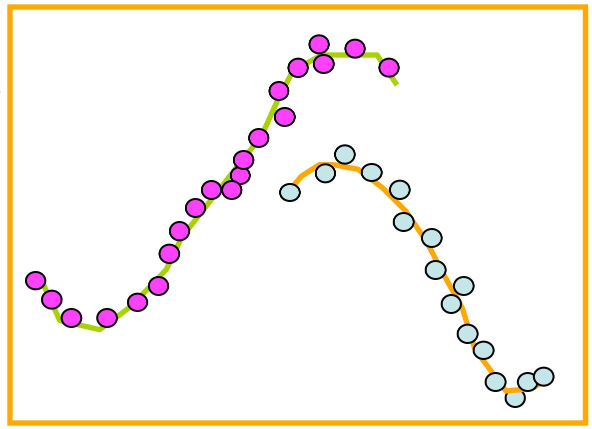
$$\sum_{i=1}^{n} \| \sum_{j=1}^{n} W_{ij}\mathbf{x}_j - \mathbf{x}_i\|^2$$

- Find $\mathbf{y}_i \in \mathbb{R}^d$ which minimize the error $\sum_{i=1}^{n} \|\mathbf{y}_i - \sum_{j=1}^{n} W_{ij}\mathbf{y}_j\|^2$
  Solve a sparse eigenvalue problem on matrix $M = (I - W)^\top (I - W)$.
  The first eigenvector is the constant vector corresponding to eigenvalue 0.

Locally linear manifold clustering

- Nonlinear manifolds
  - If the manifolds are k-separated
    - M is block-diagonal and $\dim(\text{null}(M)) = m$
    - Vectors in the null space are of the form

$$\mathbf{v}_{ij} = \begin{cases} 1 & \text{if point i belongs to group j} \\ 0 & \text{otherwise} \end{cases}$$

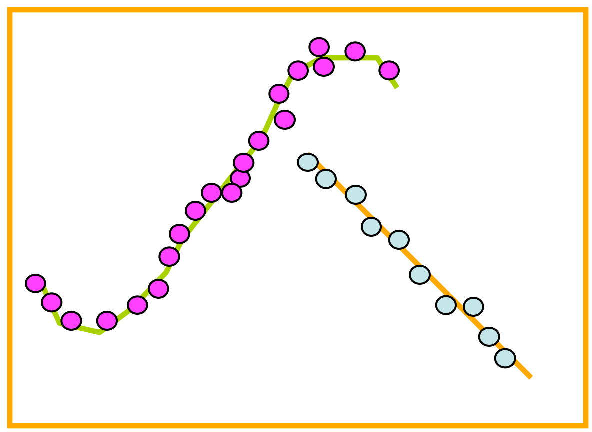


  - If the manifolds are not k-separated

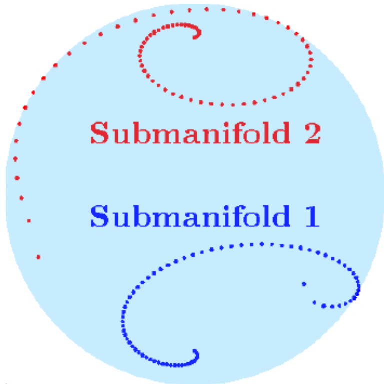
$\mathbf{v}_{ij} \approx 1$ $\mathbf{v}_{ij} \approx 0$

- Linear and nonlinear manifolds
  - $\dim(\text{null}(M)) = m + \sum d_i$
  - $M\mathbf{v} = 0$ and $M\mathbf{e} = 0$
  - If B is a basis for $\text{null}(M)$, membership vectors can be found as $\mathbf{v} = B\mathbf{x}$, where

$$\mathbf{x} = \arg\min \textstyle\sum_{ij} w_{ij}(b_i^\top \mathbf{x} - c_j)^2$$



Extending LLE to Riemannian manifolds



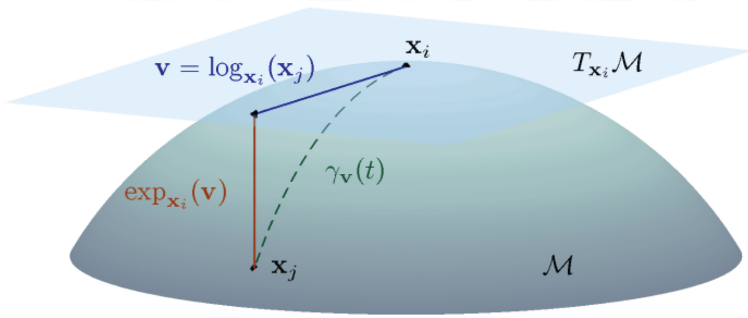


Table 1: Comparison of operations in Euclidean and Riemannian spaces.

Operation	Euclidean	Riemannian
Subtraction $\overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_j}$	$\boldsymbol{x}_j - \boldsymbol{x}_i$	$\log_{\boldsymbol{x}_i}(\boldsymbol{x}_j)$
Addition $\boldsymbol{x}_j$	$\boldsymbol{x}_i + \overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_j}$	$\exp_{\boldsymbol{x}_i}(\overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_j})$
Distance $\mathrm{dist}(\boldsymbol{x}_i, \boldsymbol{x}_j)$	$\Vert\overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_j}\Vert = \Vert\boldsymbol{x}_j - \boldsymbol{x}_i\Vert$	$\Vert\log_{\boldsymbol{x}_i}(\boldsymbol{x}_j)\Vert_{\boldsymbol{x}_i} = \sqrt{\langle\log_{\boldsymbol{x}_i}(\boldsymbol{x}_j), \log_{\boldsymbol{x}_i}(\boldsymbol{x}_j)\rangle_{\boldsymbol{x}_i}}$
Mean $\overline{\mathbf{x}}$	$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i \Rightarrow \sum_{i=1}^{n}\overrightarrow{\overline{\mathbf{x}}\boldsymbol{x}_i} = 0$	$\sum_{i=1}^{n}\log_{\overline{\mathbf{x}}}(\boldsymbol{x}_i) = 0$
Sample covariance matrix $\mathrm{cov}(\boldsymbol{x})$	$\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\mathbf{x}})(\boldsymbol{x}_i - \overline{\mathbf{x}})^\top$	$\frac{1}{n}\sum_{i=1}^{n}(\log_{\overline{\mathbf{x}}}(\boldsymbol{x}_i))(\log_{\overline{\mathbf{x}}}(\boldsymbol{x}_i))^\top$
Linear interpolation $\widehat{\boldsymbol{x}}$	$\boldsymbol{x}_i + w\overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_j}$	$\exp_{\boldsymbol{x}_i}(w\overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_j})$

- Manifold geometry essential only in first two steps of each algorithm.
  - How to select the kNN?
    - by incorporating the Riemannian distance $\Vert\log_{\mathbf{x}_i}(\mathbf{x}_j)\Vert_{\mathbf{x}_i}$
  - How to compute the matrix M representing the local geometry?

Extending LLE to Riemannian manifolds

- LLE involves writing each data point as a linear combination of its neighbors.
  - Euclidean case: need to solve a least-squares problem.
  - Riemannian case: interpolation problem on the manifold.
    - How should the data points be interpolated?

$\widehat{\boldsymbol{x}}_{Riem,i}$ is the geodesic linear interpolation of $\boldsymbol{x}_i$ by its kNN and is given by

$$\widehat{\boldsymbol{x}}_{Riem,i} = \exp_{\boldsymbol{x}_i}\Big(\sum_{j=1}^{n} W_{ij} \log_{\boldsymbol{x}_i}(\boldsymbol{x}_j)\Big).$$

    - What cost function should be minimized?

The Riemannian reconstruction error

$$\varepsilon_{Riem}(W) = \sum_{i=1}^{n} \big\| \log_{\boldsymbol{x}_i}(\widehat{\boldsymbol{x}}_{Riem,i})\big\|_{\boldsymbol{x}_i}^2 = \sum_{i=1}^{n} \Big\| \sum_{j=1}^{n} W_{ij} \log_{\boldsymbol{x}_i}(\boldsymbol{x}_j)\Big\|_{\boldsymbol{x}_i}^2.$$