Non-Isometric Manifold Learning

Analysis and an Algorithm

Piotr Dollár, Vincent Rabaud, Serge Belongie

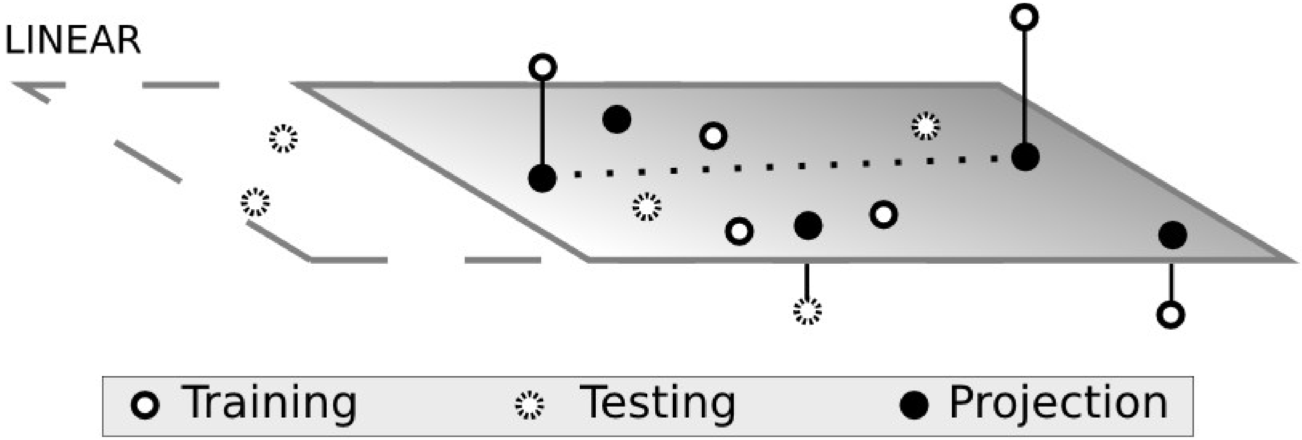
University of California, San Diego

I. Motivation

1. Extend manifold learning to applications other than embedding

2. Establish notion of test error and generalization for manifold learning

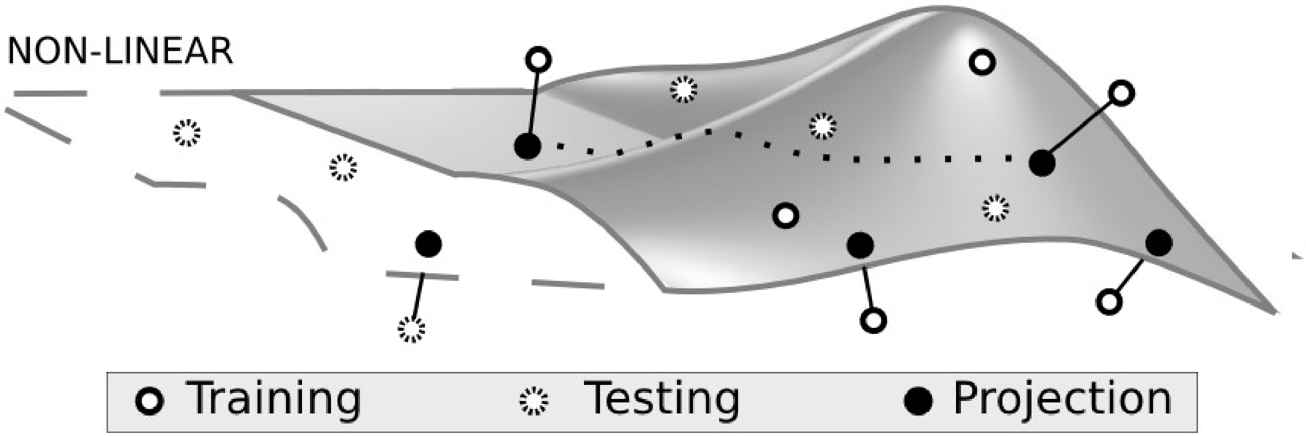
Linear Manifolds (subspaces)



Typical operations:

- project onto subspace
- distance to subspace
- distance between points
- generalize to unseen regions

Non-Linear Manifolds

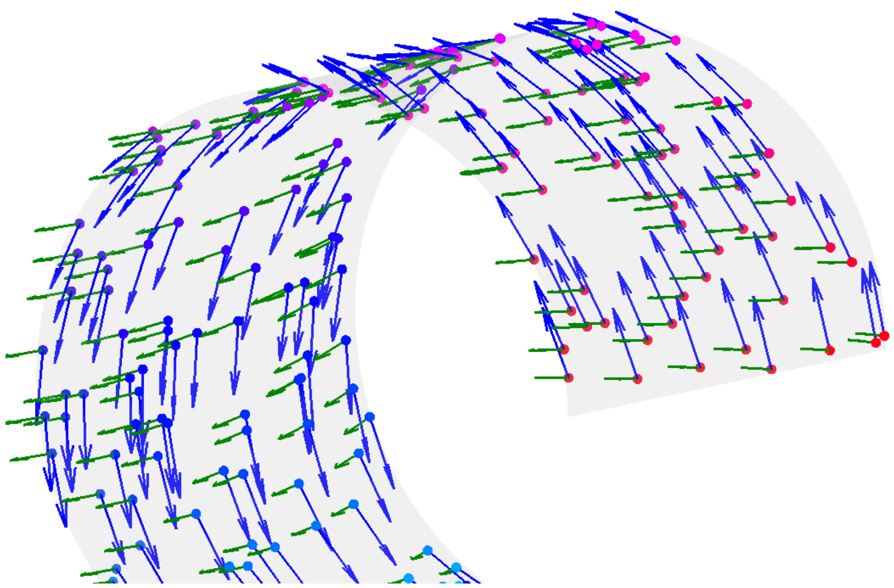


Desired operations:

- project onto manifold
- distance to manifold
- geodesic distance
- generalize to unseen regions

II. Locally Smooth Manifold Learning (LSML)

Represent manifold by its tangent space



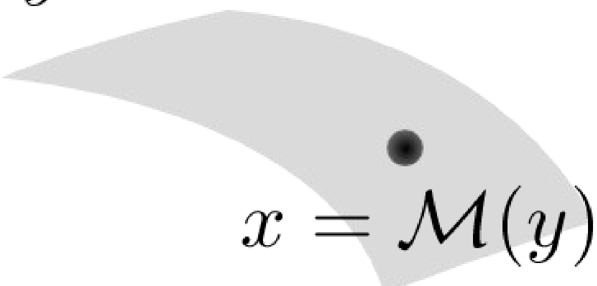
Non-local Manifold Tangent Learning [Bengio et al. NIPS05]

Learning to Traverse Image Manifolds [Dollar et al. NIPS06]

Learning the tangent space

Data on *d* dim. manifold in *D* dim. space

$y \in \mathbb{R}^d \quad x \in \mathbb{R}^D$

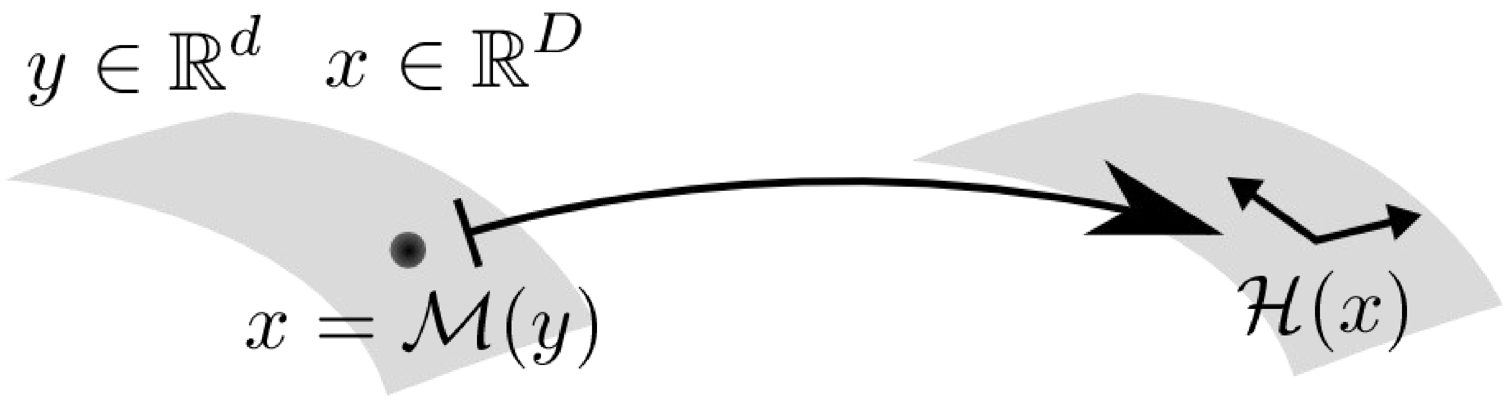


$x = \mathcal{M}(y)$

$$\mathcal{M} : \begin{cases} \mathbb{R}^d & \longrightarrow & \mathbb{R}^D \\ \mathrm{y} & \longmapsto & \mathrm{x} \end{cases}$$

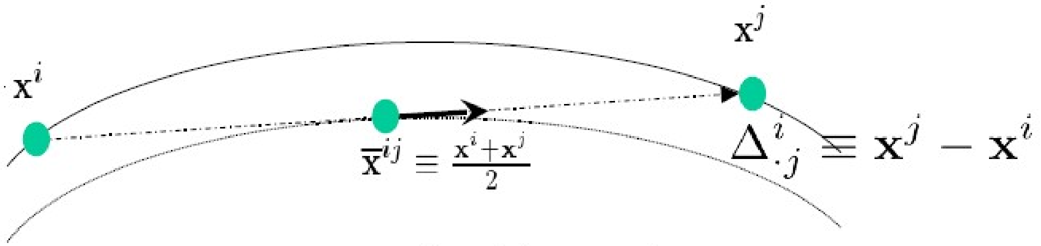
Learning the tangent space

Learn function from point to tangent basis

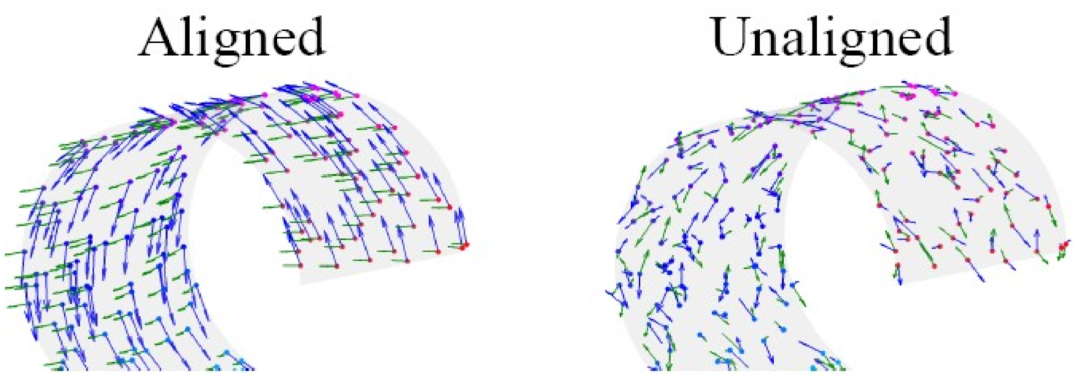


$$\mathcal{H} : \begin{cases} \mathbb{R}^D & \longrightarrow \quad \mathbb{R}^{D \times d} \\ \mathbf{x} & \longmapsto \quad \left[\frac{\partial}{\partial \mathbf{y}_1} \mathcal{M}(\mathbf{y}) \quad \cdots \quad \frac{\partial}{\partial \mathbf{y}_d} \mathcal{M}(\mathbf{y}) \right] \end{cases}$$

Loss function



$$\mathcal{H}_\theta(\overline{\mathbf{x}}^{ij})\boldsymbol{\epsilon}^{ij} \approx \Delta^i_{\cdot j}$$



$$\mathrm{err}(\theta) = \min_{\{\boldsymbol{\epsilon}^{ij}\}} \sum_{i,j \in \mathcal{N}^i} \left\| \mathcal{H}_\theta(\overline{\mathbf{x}}^{ij})\boldsymbol{\epsilon}^{ij} - \Delta^i_{\cdot j} \right\|_2^2$$

Optimization procedure

Linear form: $\mathcal{H}_\theta(\overline{\mathbf{x}}^{ij}) = \left[\Theta^1 \mathbf{f}^{ij} \cdots \Theta^D \mathbf{f}^{ij}\right]^\top$

$$\mathrm{err}(\theta) = \min_{\{\epsilon^{ij}\}} \sum_{i,j \in \mathcal{N}^i} \sum_{k=1}^{D} \left(\mathbf{f}^{ij\top} \Theta^{k\top} \boldsymbol{\epsilon}^{ij} - \Delta^i_{kj} \right)^2$$

Initialize Θ randomly.

while $\mathrm{err}(\Theta)$ decreases **do**

$\forall i, j$, solve for the best $\boldsymbol{\epsilon}^{ij}$ given the Θ^ks:

$$\boldsymbol{\epsilon}^{ij} = (H^{ij\top} H^{ij} + \lambda_\epsilon I)^+ H^{ij\top} \Delta^i_{\cdot j}$$

$\forall k$, solve for the best Θ^k given the $\boldsymbol{\epsilon}^{ij}$'s:

$$\text{Let: } A = \begin{bmatrix} \vdots \\ \boldsymbol{\epsilon}^{ij\top} \otimes \mathbf{f}^{ij\top} \\ \vdots \end{bmatrix}, \mathbf{b}^k = \begin{bmatrix} \vdots \\ \Delta^i_{kj} \\ \vdots \end{bmatrix}$$

$$\mathrm{vec}\left(\Theta^{k\top}\right) = (A^\top A + \lambda_\Theta (I \otimes (\Delta_F{}^\top \Delta_F))^{-1} A^\top \mathbf{b}^k$$

end while

III. Analyzing Manifold Learning Methods

- Need evaluation methodology to:
  - objectively compare methods
  - control model complexity
  - extend to non-isometric manifolds

Evaluation metric

By definition, for isometric manifolds embedding should preserve distance

$$\text{err}_{\text{GD}} \equiv \frac{1}{n^2} \sum_{ij} \frac{|d_{ij} - d'_{ij}|}{d_{ij}}$$

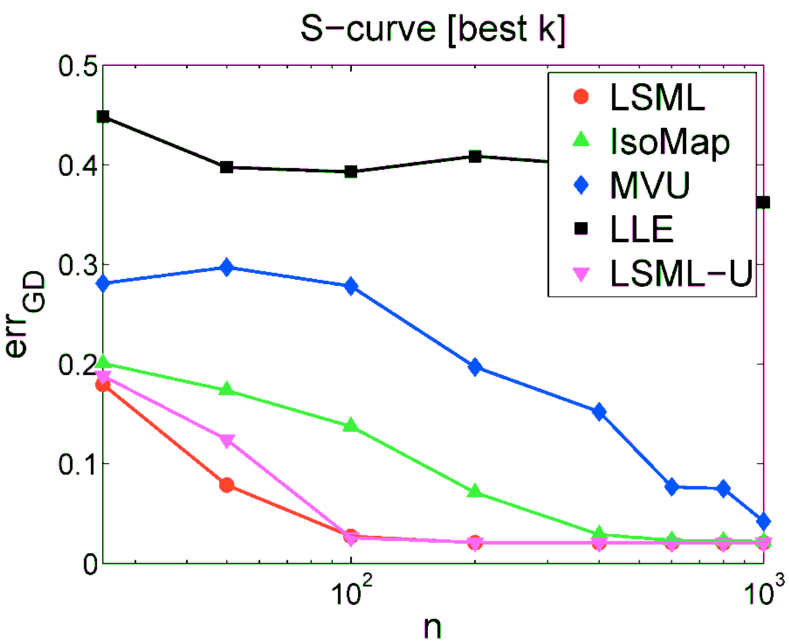
(d'_{ij} ← Estimated dist; d_{ij} ← True distance)

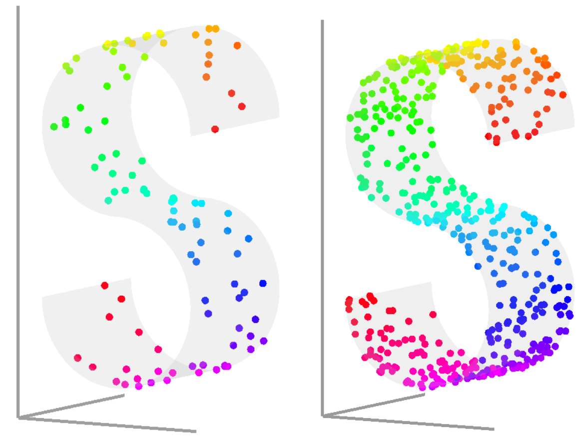
Requires two sets of samples from manifold:

1. S_n – for training
2. S_∞ – for computing "true" geodesics (Isomap)

Applicable for manifolds that can be densely sampled

Finite Sample Performance

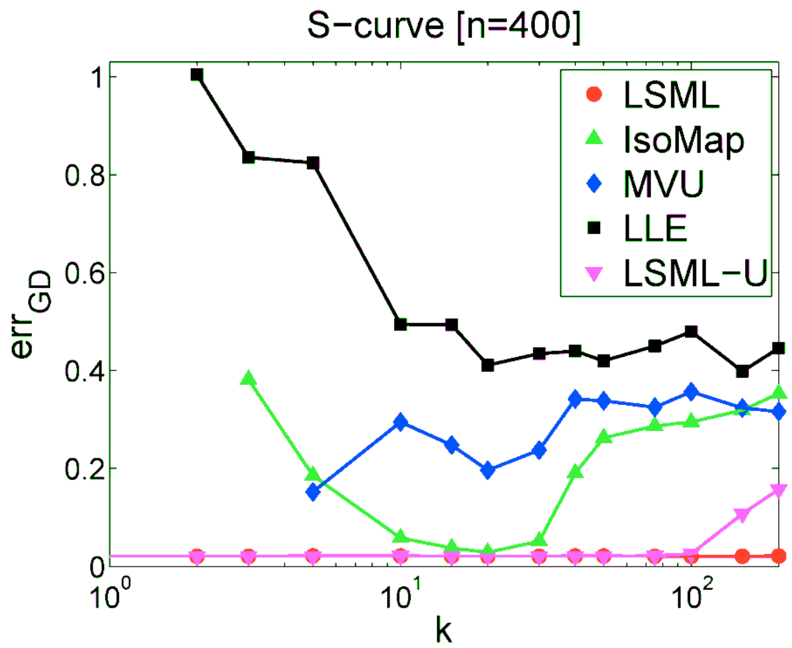


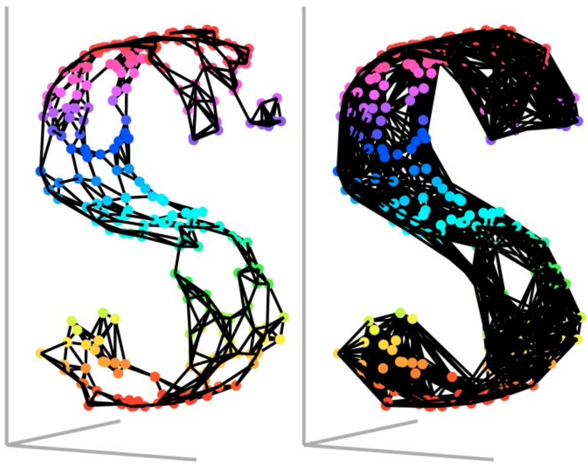


Performance: LSML > ISOMAP > MVU >> LLE

Applicability: LSML >> LLE > MVU > ISOMAP

Model Complexity





All methods have at least one parameter: k

Bias-Variance tradeoff

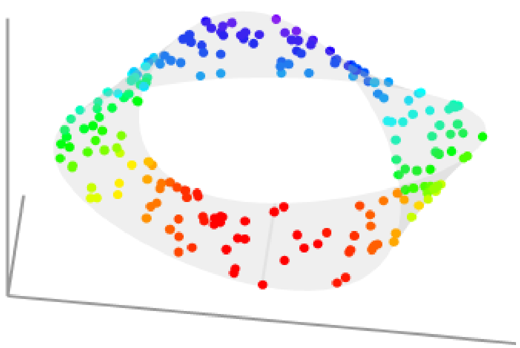
LSML test error

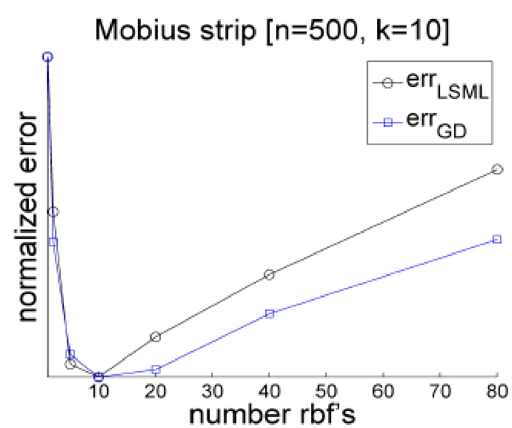
Typically, S_∞ not available

- Need notion of generalization: **testable prediction**
- Model assessment / Model selection

Define: $$\mathrm{err}_{\mathrm{LSML}} \equiv \sum_i \min_{\boldsymbol{\epsilon}^{ii'}} \left\| \mathcal{H}_\theta(\overline{\mathbf{x}}^{ii'})\boldsymbol{\epsilon}^{ii'} - (\mathbf{x}^i - \mathbf{x}^{i'}) \right\|_2^2$$

Claim: $\mathrm{err}_{\mathrm{LSML}}$ / $\mathrm{err}_{\mathrm{GD}}$ strongly correlated:





LSML test error

Typically, S_∞ not available

- Need notion of generalization: **testable prediction**
- Model assessment / Model selection

Define: $$\mathrm{err}_{\mathrm{LSML}} \equiv \sum_i \min_{\boldsymbol{\epsilon}^{ii'}} \left\| \mathcal{H}_\theta(\overline{\mathbf{x}}^{ii'})\boldsymbol{\epsilon}^{ii'} - (\mathbf{x}^i - \mathbf{x}^{i'}) \right\|_2^2$$

- Use much as test error in supervised learning
- Cannot be used to select *d*
- Can also measure error for manifold transfer

IV. Using the Tangent Space

- projection
- manifold de-noising
- geodesic distance
- generalization

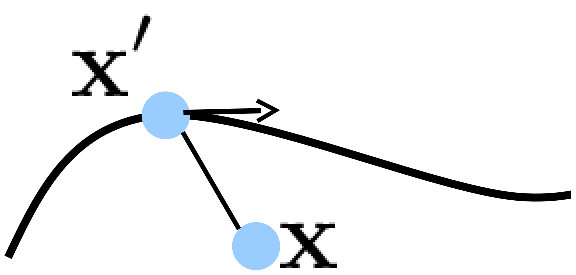
Projection

x' is the projection of *x* onto a manifold if it satisfies:

$$\min_{\mathbf{x}'} \|\mathbf{x} - \mathbf{x}'\|_2^2$$

gradient descent is performed after initializing the projection *x*' to some point on the manifold:

$$\mathbf{x}' \leftarrow \mathbf{x}' + \alpha H' H'^{\top} (\mathbf{x} - \mathbf{x}')$$



$H' \longleftarrow$ tangents at $\mathbf{x}'$

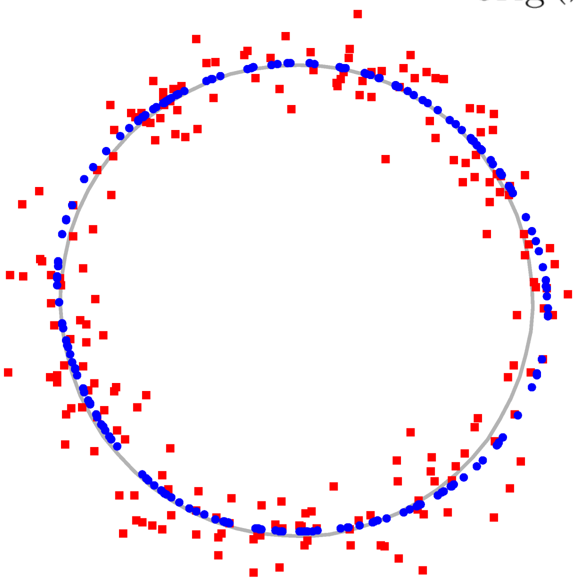
$H' H'^{\top} \longleftarrow$ projection matrix

Manifold de-noising

Goal: recover points on manifold (χ) from points corrupted by noise ($\mathbf{X}$)

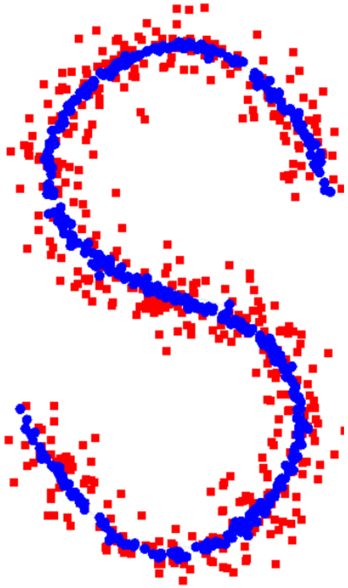
$$\mathrm{err}_{\mathcal{M}}(\chi) = \min_{\{\boldsymbol{\epsilon}^{ij}\}} \sum_{i,j \in \mathcal{N}^i} \left\| \mathcal{H}_\theta(\overline{\chi}^{ij})\boldsymbol{\epsilon}^{ij} - (\chi^i - \chi^j) \right\|_2^2$$

$$\mathrm{err}_{\mathrm{orig}}(\chi) = \sum_{i=1}^{n} \left\| \chi^i - \mathbf{x}^i \right\|_2^2$$



$\mathbf{X}$ - original

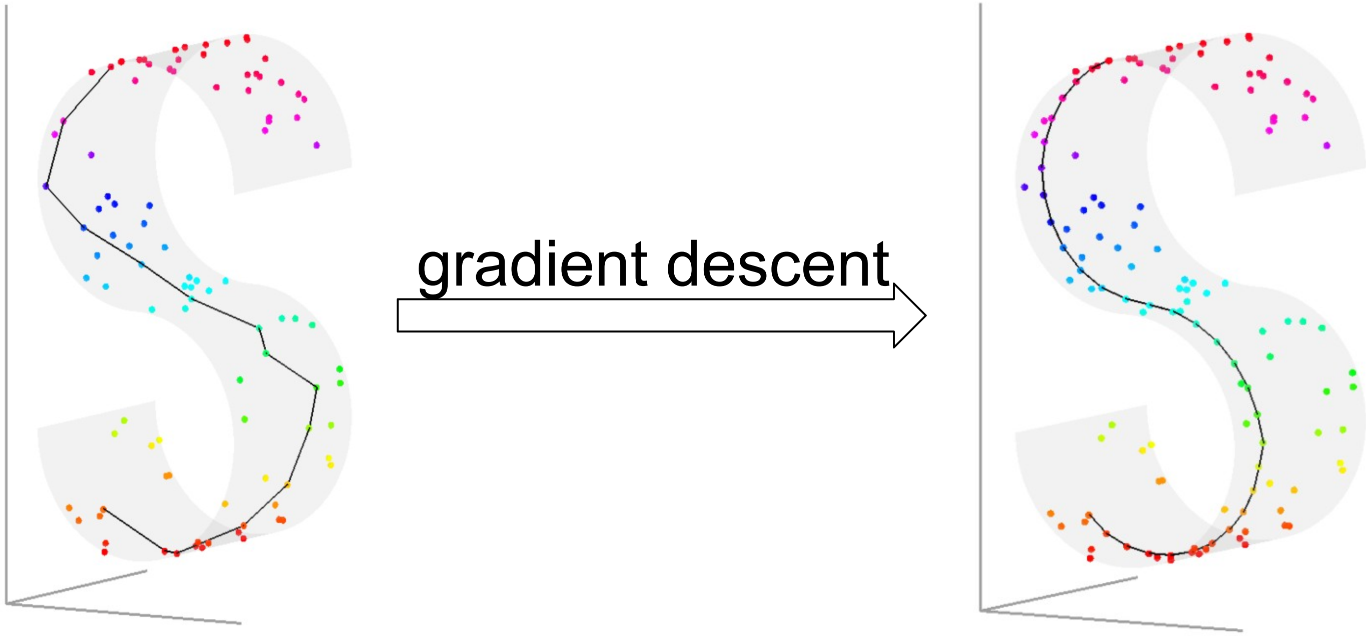
χ - de-noised



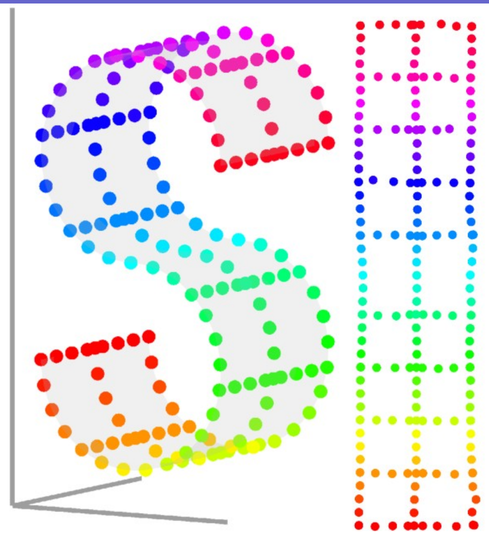
Geodesic distance

Shortest path: $$\mathrm{err}_{\mathrm{length}}(\chi) = \sum_{i=2}^{m} \left\| \chi^i - \chi^{i-1} \right\|_2^2$$

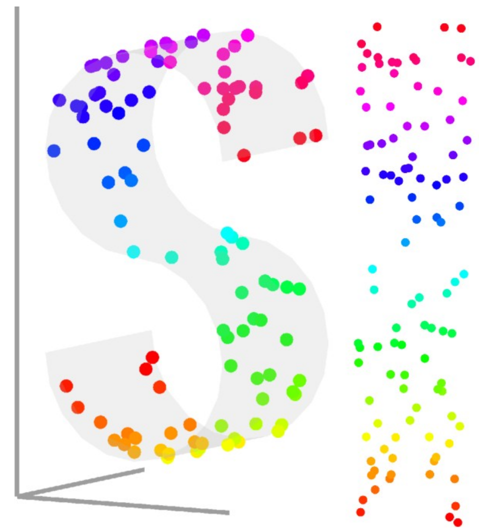
On manifold: $$\mathrm{err}_{\mathcal{M}}(\chi) = \min_{\{\boldsymbol{\epsilon}^{ij}\}} \sum_{i,j \in \mathcal{N}^i} \left\| \mathcal{H}_\theta(\overline{\chi}^{ij})\boldsymbol{\epsilon}^{ij} - (\chi^i - \chi^j) \right\|_2^2$$

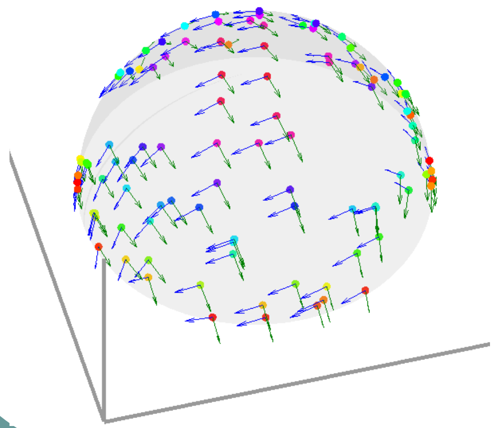


Geodesic distance

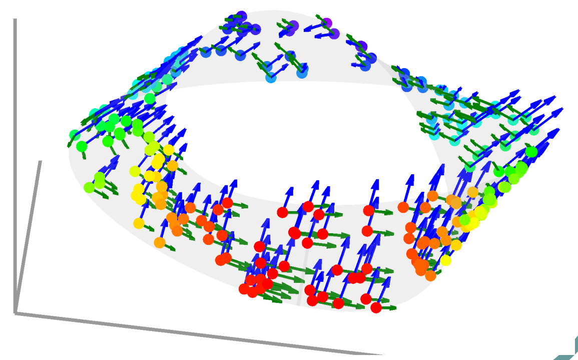


← embedding of sparse and structured data →

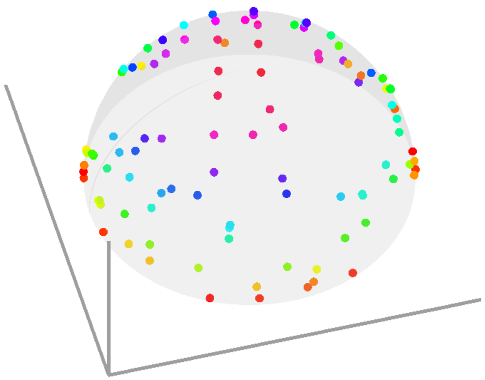


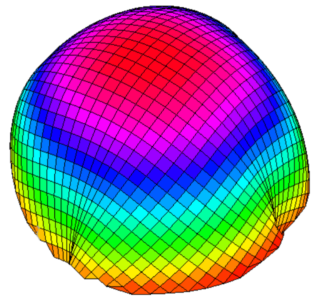


←can apply to non-isometric manifolds →

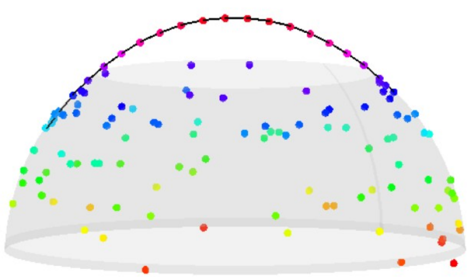


Generalization





← Reconstruction within the support of training data



← Generalization beyond support of training data →

Thank you!