Learning a manifold from a teacher's demonstrations

By Pei Wang, Arash Givchi and Patrick Shafto

RUTGERS
UNIVERSITY | NEWARK
CoDaS Lab

1. Goal: establish a manifold learning approach based on structured data

A main task in learning a manifold is to infer its correct topology. Inspired by human learners, we:

✓ propose two methods to teach the topology of a manifold: **isolated data points (individual examples)** and **sequential data points (demonstrations)**.

✓ provide theoretical **bound on the minimal number of points** required for learning a manifold for each method.

✓ demonstrate that the approach based on **structured data requires a significantly smaller minimum number of data points.**

2. Formalism of teaching-learning algorithms

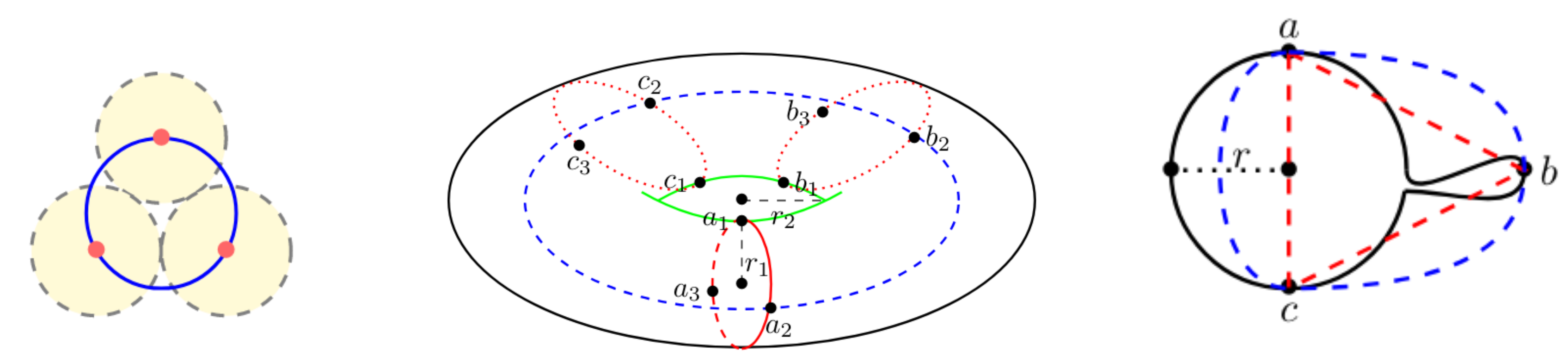
- **Learning algorithm**: a class of algorithm $\mathscr{A}$ that identify the homotopy type of a manifold $\mathscr{M}$ from a set of data sampled from $\mathscr{M}$.
- **Teaching set**: a collection of data $\mathscr{D}_{\mathscr{A}} \subset \mathscr{M}$ that can be used to recover the homotopy type of $\mathscr{M}$ from at least a learning algorithm $A \subset \mathscr{A}$.
- **Minimal teaching number**: a teaching set $\mathscr{D}^*_{\mathscr{A}}$ is said to be *minimal w.r.t. $\mathscr{M}$ (or homotopy type of $\mathscr{M}$)* if it contains the least number of data points among all teaching sets. The size of $\mathscr{D}^*_{\mathscr{A}}$ is called the *minimal teaching number.*

3. Manifold teaching from sample points

Rephrasing pioneering work by *Niyogi et al.[2008],* we have:

- **Teacher:** pass randomly sampled isolated data points $\mathscr{D}_{\mathscr{A}(\epsilon)}$;
- **Learner:** pick an $\epsilon \in \mathbb{R}^+$, then for each $x \in \mathscr{D}_{\mathscr{A}(\epsilon)}$, makes an n-dim ball $B_\epsilon(x)$ centered at x of radius ϵ. The union of these balls constitutes the learned space.

✤ **Example 1:** a unit circle can be taught by three equidistant sampled points as shown below:



✤ **Example 2:** 9 points showed above is not enough to teach a regular torus. Detail calculation shows that at least 51 points are needed.

Proposition 3.3: Let $\mathscr{M}_g \in \mathbb{R}^3$ be a closed orientable surface with genus g. Then the minimal teaching number for the homotopy type of $\mathscr{M}_g$ with respect to $\mathscr{A}(\epsilon)$ is bounded by $49g + 2$.

5. Discussion and Further work

- **Question 1:** What does the teacher need to know in order to choose a sequence for demonstration?

  Intuition: sufficient data is required to estimate the correct topology in the first place $\Longrightarrow$ ***topological reach.***

- **Question 2:** Would learning the topology of a manifold accelerate the learning of geometry afterwards?
- **Question 3:** What makes a class of algorithms good? and how to quantify such goodness?

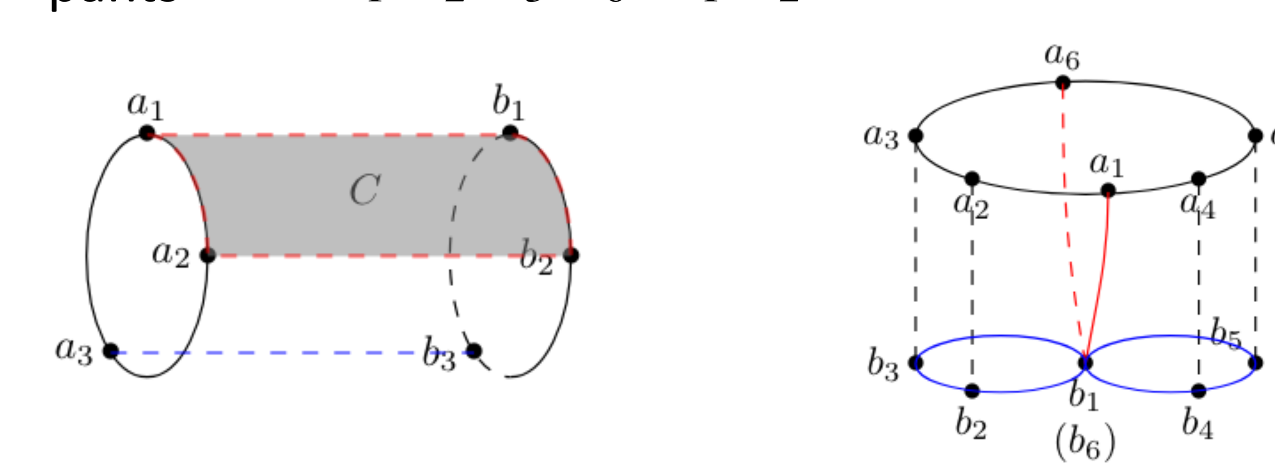
4. Manifold teaching from demonstrations

- **Problem:** the locally Euclidean nature of manifolds forces large amount of data for any local to global teaching procedure.
- **Solution:** we propose a method that teaches the topology directly from demonstrations where each demonstration is a sequence of data describing a loop.
  - **Teacher:** pass a set of sequential data $\mathscr{D}_{\mathscr{A}(\mathbf{l})}$;
  - **Learner:** pick local structure with proper dimension to connect consecutive sequential data. The obtained space $U_{\mathscr{D}_{\mathscr{A}(\mathbf{l})}}$ is the learned space.

✤ **Example 3:** the black loop shown in the figure on the left can be taught by $\mathscr{D}_{\mathscr{A}(\mathbf{l})} = \{[a, b, c, d]\}$.
The sequential data informs the learner to connect consecutive points by simple curves. The red and blue curves are two examples of $U_{\mathscr{D}_{\mathscr{A}(\mathbf{l})}}$.

✓ Any (oriented) non-contractible 1-dim manifold can be effectively described by a sequence consisting three randomly sampled points.

✤ **Example 4:**
* **A torus** can be taught by
$$\mathscr{D}_{\text{torus}} = [[a_1, a_2, a_3, a_1][b_1, b_2, b_3, b_1][c_1, c_2, c_3, c_1][a_1, a_2, a_3, a_1]]$$
* **A cylinder** can be taught by :
$$\mathscr{D}_{\text{cylinder}} = \{[[a_1, a_2, a_3, a_4], [b_1, b_2, b_3, b_1]]\}.$$
***A pair of pants** can be taught by
$$\mathscr{D}_{\text{pants}} = \{[[a_1, a_2, a_3, a_6][b_1, b_2, b_3, b_6]], [[a_1, a_4, a_5, a_6][b_1, b_4, b_5, b_6]]\}$$



Proposition 3.7: Let $\mathscr{M}_g \in \mathbb{R}^3$ be a closed orientable surface with genus $g \geq 2$. Then the minimal teaching number for of $\mathscr{M}_g$ with respect to $\mathscr{A}(\mathbf{l})$ is bounded by $3g - 3$ sequences where each sequence consists of at most 4 data points.