Summer@ICERM 2017: Topological Data Analysis
Mini-course on *Distances between metric spaces and applications*

Lecturer: Facundo Mémoli

TA: Samir Chowdhury

Brief description. This course is an introduction to Metric Geometry and applications. Of particular importance will be the definition and properties of the so called Gromov-Hausdorff distance between metric spaces and applications to shape and data analysis and matching. We'll also look into the setting of directed/asymmetric networks.

Introduction. Finite metric spaces are a natural model for data. From a finite metric space (X, d_X) one can induce several different simplicial filtrations. Examples include the Rips and Witness filtrations. From these simplicial filtrations, via the mechanism of persistent homology one obtains barcodes (or persistence diagrams) $\mathcal{B}(X)$ as a summary of the metric information contained in the data.

A main goal of this mini-course will be to establish the mathematical and computational language to express the stability of the assignment $X \mapsto \mathcal{B}(X)$. In order to do this, one the one hand one needs a way of measuring distance between the "input": that is, one needs a gadget that takes two finite metric spaces (X, d_X) and (Y, d_Y) and produces a number measuring the distance between these objects. On the other hand one needs a notion of distance between the outputs: the barcodes $\mathcal{B}(X)$ and $\mathcal{B}(Y)$.

The notion of distance that we'll use for measuring the discrepancy between $\mathcal{B}(X)$ and $\mathcal{B}(Y)$ will be the so called bottleneck distance $d_B(\mathcal{B}(X), \mathcal{B}(Y))$. And, in order to measure distance between finite metric spaces we'll use the Gromov-Hausdorff distance $d_{GH}(X, Y)$. In particular, we'll eventually proving the following theorem:

Theorem 1. *Let (X, d_X) and (Y, d_Y) be any two finite metric spaces and let k be any non-negative integer. Let $\mathcal{B}^R_k(X)$ and $\mathcal{B}^R_k(Y)$ be the respective Rips Homology barcodes in dimension k. Then,*

$$2\, d_{GH}(X, Y) \geq d_B(\mathcal{B}^R_k(X), \mathcal{B}^R_k(Y)). \tag{1}$$

Now, a fact is that the computation of the GH distance between finite metric spaces leads to a very "complex" algorithmic problem. In fact, theoretical comuter scientist refer to a problem of this type as NP-hard. In contrast, the right hand side of (1) can be computed in time which depends polynomially on the "size" of the inputs X and Y. This means that the lower bound provided by the theorem gives us a computationally tractable way of estimating the distance between finite metric spaces (datasets). This, as we will see in this course, has a strong impact on applications.

Example 1. *For example, pick a real number $\epsilon > 0$ much smaller than 1. Then in the above theorem let $k = 0$, X be the metric space with two points at distance one from each-other, and let Y be the metric space with two points at distance $(1+\epsilon)$ from each-other, then $d_{GH}(X, Y) = \frac{\epsilon}{2}$. On the other hand, $\mathcal{B}^R_k(X) = \{[0, 1), [0, \infty)\}$ and $\mathcal{B}^R_k(Y) = \{[0, 1+\epsilon), [0, \infty)\}$. Then, one can verify that the bottleneck distance is ϵ. This means that these choices for X and Y produce an equality in the bound given in the theorem, therefore, the ound is* tight.

In order to prove the theorem above we'll have to study a number of additional tools, including the notion of the *interleaving distance* between persistence modules.

Format: The course consists of a blend of theoretical and practical/computational sessions. During the computational sessions we'll be running examples in Matlab. Some of these examples will build upon the JavaPlex software you will learn in Henry's course: `http://appliedtopology.github.io/javaplex/`. Javaplex tutorial: `http://www.math.colostate.edu/~adams/research/javaplex_tutorial.pdf`

Reading materials
This mini-course will rely heavily on the material of the first week (simplicial complexes, filtrations, persistent homology of finite metric spaces). Besides basic/minimal knowledge about metric spaces, students will be expected to have a working knowledge of the persistent homology of Rips filtrations, the bottleneck distance between persistence diagrams, as well as familiarity with software tools (e.g., javaPlex) to carry out explicit computations.

Assignment 1 (Reading assignment)**.** *Start reading [1] and complement with material from [3] as needed.*

General references are included below.

References

[1] Frédéric Chazal, David Cohen-Steiner, Leonidas J Guibas, Facundo Mémoli, and Steve Y Oudot. *Gromov-hausdorff stable signatures for shapes using persistence.* In *Computer Graphics Forum*, volume 28, pages 1393–1403. Wiley Online Library, 2009.

[2] Facundo Mémoli. *Some Properties of Gromov–Hausdorff Distances." Discrete & Computational Geometry* 48.2 (2012): 416-440.

[3] A course on metric geometry. Burago, Burago, Ivanov.
`http://bookstore.ams.org/gsm-33`

[4] Robert Ghrist. Elementary Applied Topology. `https://www.math.upenn.edu/~ghrist/notes.html`.

[5] Gunnar Carlsson. *Topological pattern recognition for point cloud data. Acta Numerica*, 23 (2013): 289–368.

[6] Herbert Edelsbrunner and John Harer. *Computational topology: An introduction*. American Mathematical Society, 2010.

[7] `http://tosca.cs.technion.ac.il/book/resources_data.html`

[8] Samir Chowdhury and Facundo Mémoli. *Persistent Homology of Asymmetric Networks: An Approach based on Dowker Filtrations.*

`https://arxiv.org/abs/1608.05432`