INF556 Topological Data Analysis
Final Exam — 3 hours

December 22, 2017

Important:

- The exercises and problem are independent of one another. **Please provide your solution to the problem on a separate sheet**.
- The text of the exam is long on purpose. This is to ensure that everyone can at least solve some problem(s). **Please keep in mind that the quality of your solutions is more important than their quantity** (to some extent, of course).
- The text of the exam is written in English. Your answers can be written indifferently in French or in English.
- All printed documents are allowed. By contrast, computers, cellphones, tablets, pocket calculators, smart watches, etc., are forbidden.

1 PCA vs MDS

Let $P = \{p_1, \cdots, p_n\}$ be a point cloud in $\mathbb{R}^d$. Let $M \in \mathbb{R}^{d\times n}$ be its coordinates matrix, which has one column per data point and one line per dimension. Suppose that P is centered, that is:

$$M \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Let us fix a target dimension $k \leq d$ and apply PCA and MDS on M.

Question 1. Show that PCA and MDS output the same $k \times n$ matrix (called K hereafter).
Hint: use the Singular Value Decomposition of M.

Question 2. Give a procedure to compute K without building the covariance matrix MM^T or Gram matrix M^TM.

2 Scissors and glue...

Question 3. The projective line $\mathbb{RP}^1$ is obtained from the unit circle $\mathbb{S}^1$ by identifying antipodal points: $x \sim -x$ for every $x \in \mathbb{S}^1$. Show that $\mathbb{RP}^1$ is homeomorphic to the circle $\mathbb{S}^1$ itself. A proof by pictures showing the sequence of gluing and twisting operations will be enough.

Question 4. Show that the quotient space depicted in Figure 1 is homeomorphic to the torus. Again, a proof by pictures showing the sequence of cuttings and gluings will be enough.

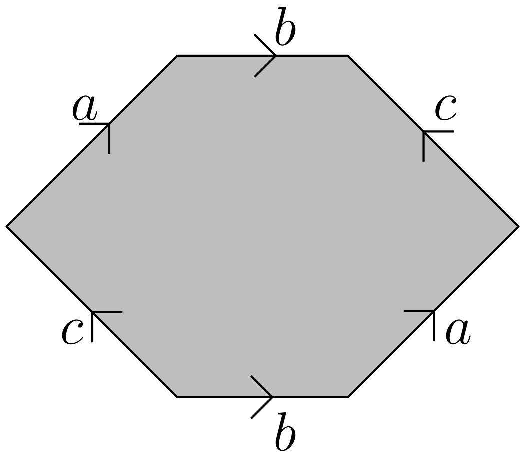

Figure 1: The opposite sides of the hexagon are glued together with the given orientations.

3 Some calculations...

Let the field of coefficients be $\mathbb{Z}/2\mathbb{Z}$.

Question 5. Compute the barcode associated with the simplicial filtration depicted in Figure 2.

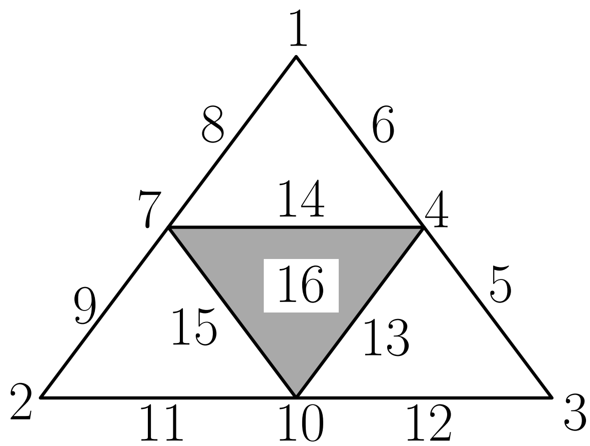

Figure 2: A simplicial filtration.

4 Euler characteristic

Given a triangulable topological space X and a field $\mathbf{k}$, the *Euler characteristic* is the quantity:

$$\chi(X;\mathbf{k}) = \sum_{i=0}^{+\infty}(-1)^i \dim \mathrm{H}_i(X;\mathbf{k}).$$

Note that this quantity is finite because each homology group is finite-dimensional and only finitely many of them are non-trivial, X being triangulable.

Question 6. Show that χ is a topological invariant, that is: for any spaces X, Y that are homotopy equivalent, $\chi(X; \mathbf{k}) = \chi(Y; \mathbf{k})$.
Hint: look at what happens to each individual homology group.

Now we want to prove the Euler-Poincaré theorem:

Theorem 1. *For any simplicial complex X and any field $\mathbf{k}$:*

$$\chi(X; \mathbf{k}) = \sum_{i=0}^{+\infty} (-1)^i n_i(X),$$

where $n_i(X)$ denotes the number of simplices of X of dimension i.

For this we will use topological persistence. Consider an arbitrary filtration of X:

$$\emptyset = X_0 \subseteq X_1 \subseteq \cdots \subseteq X_m = X.$$

Assume without loss of generality that a single simplex σ_j is inserted at each step j:

$$\forall j = 1, \cdots, m, \; X_j \setminus X_{j-1} = \{\sigma_j\}.$$

Note that m is then equal to the number of simplices of X, that is:

$$m = \sum_{i=0}^{+\infty} n_i(X).$$

Let us apply the persistence algorithm to this simplicial filtration. Recall from lecture 5 that we have the following property:

Lemma 2. *At each step j, the insertion of simplex σ_j either creates an independent d_j-dimensional cycle (i.e. increases the dimension of $\mathrm{H}_{d_j}(X_{j-1}; \mathbf{k})$ by 1) or kills a $(d_j - 1)$-dimensional cycle (i.e. decreases the dimension of $\mathrm{H}_{d_j-1}(X_{j-1}; \mathbf{k})$ by 1), where d_j is the dimension of σ_j.*

Question 7. Using Lemma 2, prove Theorem 1.
Hint: proceed by induction on j.

Question 8. Deduce that the Euler characteristic of a triangulable space is independent of the choice of field $\mathbf{k}$.

5 The Dunce Hat

Recall that the Dunce Hat is obtained by indentifying the three edges of a triangle as shown in Figure 3.

Question 9. Build a triangulation of the Dunce Hat (you may draw a picture to represent it). Beware that your triangulation must be a simplicial complex.

Question 10. Use your simplicial complex to compute the homology of the Dunce Hat.
Hint: to avoid tedious calculations, you can proceed as in exercise 4: pick a filtration of your complex then apply the persistence algorithm; for each simplex σ_j inserted, use Lemma 2 to predict its effect on the homology (identify the created d_j-cycle or the killed $(d_j - 1)$-cycle).

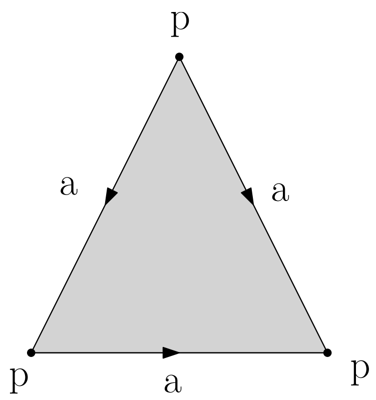

Figure 3: The Dunce Hat.

6 Decomposition of persistence modules

Let us fix a **finite** index set $T \subseteq \mathbb{R}$ of cardinality $\#T < +\infty$, and an arbitrary field of coefficients, and let us consider **pointwise finite-dimensional** persistence modules over T, i.e. persistence modules $\mathbb{V}$ such that $\dim V_t < +\infty$ for all $t \in T$.

A *submodule* $\mathbb{W}$ of a persistence module $\mathbb{V}$ is composed of subspaces $W_t \subseteq V_t$ for all $t \in T$, and of maps $w_t^{t'} = v_t^{t'}|_{W_t}$ for all $t \leq t' \in T$. In particular, $v_t^{t'}(W_t) \subseteq W_{t'}$ for all $t \leq t' \in T$. A simple example of submodule is the *null module* $\mathbb{W} = 0$ (defined by $W_t = 0$ for all $t \in T$ and $w_t^{t'} = 0$ for all $t \leq t' \in T$), which is a submodule of any module $\mathbb{V}$ over T.

Direct sums in the category are defined pointwise, that is: for any persistence modules $\mathbb{V}$ and $\mathbb{W}$ over T, the direct sum $\mathbb{V} \oplus \mathbb{W}$ is composed of the spaces $V_t \oplus W_t$ for all $t \in T$, and of the maps $v_t^{t'} \oplus w_t^{t'}$ for all $t \leq t' \in T$. We say that $\mathbb{V}$ and $\mathbb{W}$ are *summands* of the direct sum $\mathbb{V} \oplus \mathbb{W}$. In particular, we have $\mathbb{V} = 0 \oplus \mathbb{V} = \mathbb{V} \oplus 0$ for any persistence module $\mathbb{V}$, so $\mathbb{V}$ is always a summand of itself. A persistence module $\mathbb{V}$ is called *indecomposable* if its only summands are itself or the null module. The decomposition theorem that we saw in class asserts that the only indecomposable persistence modules are the so-called *interval modules* over T, and that (under some conditions on T or on the dimensions of the spaces) any module decomposes as a direct sum of interval modules.

Question 11. In the case where $\#T = 1$, show that every submodule is a summand. Deduce the decomposition theorem in the case $\#T = 1$.

In the general case, the result is more complicated to prove as submodules may not always be summands:

Question 12. In the case where $\#T \geq 2$, exhibit a counterexample showing that a submodule of a persistence module $\mathbb{V}$ over T may not always be a summand of $\mathbb{V}$.

The case $\#T = 2$ is somewhat simpler to handle though, as only one linear map is involved.

Question 13. Prove the decomposition theorem in the case where $\#T = 2$.

The general (finite) case involves more technicalities. First, you must show that your module actually decomposes into indecomposables:

Question 14. Show that every (pointwise finite-dimensional) module over T (finite) decomposes into indecomposable modules. You may proceed for instance by induction on the *total dimension* of $\mathbb{V}$, defined as the quantity $\sum_{t \in T} \dim V_t$.

Then, the structure of each indecomposable must be studied:

Question 15 (difficult). Show that every indecomposable (pointwise finite-dimensional) module over T (finite) is an interval module.

The decomposition theorem follows. Note that we have not considered the uniqueness of the decomposition here.

7 Problem: the space of persistence diagrams

Important reminder: you must write your solution to this problem on a separate sheet.

7.1 Introduction

Let $(\mathcal{X}, d_{\mathcal{X}})$ be a metric space. For $x, y \in \mathcal{X}$, we call a *path* from x to y a continuous map $\gamma : [0,1] \to X$ such that $\gamma(0) = x, \gamma(1) = y$. We note $\Gamma(x,y)$ the set of all such paths. For all the remainder of this problem, we will suppose that $\mathcal{X}$ is always *path-connected*, that is $\Gamma(x,y)$ is never empty for any $x, y \in \mathcal{X}$.

A *subdivision* of $[0,1]$ is given by an integer $n \geqslant 2$ and a n-tuple $t_1 \leqslant .. \leqslant t_n$ such that $t_1 = 0, t_n = 1$. We write $\mathcal{S}$ the set of all subdivisions. We define the *length* of a path γ to be:

$$L(\gamma) := \sup_{(t_1..t_n)\in\mathcal{S}} \sum_{i=1}^{n-1} d_{\mathcal{X}}(\gamma(t_i), \gamma(t_{i+1})) \tag{1}$$

We now define the *geodesic distance* on $\mathcal{X}$ to be:

$$d_g(x,y) := \inf_{\gamma\in\Gamma(x,y)} L(\gamma) \tag{2}$$

We finally say that a path $\gamma \in \Gamma(x,y)$ is a *geodesic* between x and y if it achieves the infimum in (2), and that $(\mathcal{X}, d_{\mathcal{X}})$ is a *geodesic space* if any two points x, y are connected by such a geodesic (i.e. the infimum in (2) is always achieved).

7.2 General facts

Question 16. Let $(\mathcal{X}, d_{\mathcal{X}})$ be a metric space. Show that $\forall x, y \in \mathcal{X}, d_g(x,y) \geqslant d_{\mathcal{X}}(x,y)$

Question 17. Show that $(\mathbb{R}^d, ||.||_2)$ is a geodesic space.

Question 18. Let $(\mathcal{X}, d_{\mathcal{X}})$ be a geodesic space. Show that the geodesic distance introduced in (2) is indeed a distance over $\mathcal{X}$, i.e it satisfies the following axioms:

- $\forall x, y \in \mathcal{X}, d_g(x,y) = 0 \Leftrightarrow x = y$
- $\forall x, y \in \mathcal{X}, d_g(x,y) = d_g(y,x)$
- $\forall x, y, z \in \mathcal{X}, d_g(x,z) \leqslant d_g(x,y) + d_g(y,z)$

7.3 Geodesics for persistence diagrams

In the following, we consider that a persistence diagram is a **finite** multiset of points[1] in $\mathbb{R}^2_> := \{(x_1, x_2) \in \mathbb{R}^2 | x_2 > x_1\}$ (called the *off-diagonal points*), together with a unique virtual point (representing the diagonal) $\{\Delta\}$ with infinite (countable) multiplicity. We say that ϕ is a *matching* between X and Y, noted $\phi \in \Pi(X,Y)$, if ϕ is a bijection between the points in X (including all copies of $\{\Delta\}$) and the points in Y (including all copies of $\{\Delta\}$).

[1] A multiset is a set of points with multiplicities. Given a point x, its copies can be labeled e.g. $x^{(1)}..x^{(n)}$ and treated as different points despite their being located at the same place.

Remark: This is equivalent to the definition given during the lectures but will be more convenient to use here.

We endow the space of persistence diagrams with the d_2 metric: for two diagrams X, Y we have,

$$d_2(X,Y) = \left(\inf_{\phi \in \Pi(X,Y)} \sum_{x \in X} ||x - \phi(x)||_2^2 \right)^{\frac{1}{2}} \tag{3}$$

with the convention that $||x - \Delta||_2 = ||x - \pi_\Delta(x)||_2$, where $\pi_\Delta(x)$ denotes the orthogonal projection of x onto the diagonal, and $||\Delta - \Delta||_2 = 0$.

A matching ϕ between X and Y which minimizes (3) is said to be *optimal.*

Goal: The goal of this subsection is to show that the set of persistence diagrams endowed with the d_2 metric is a geodesic space.

Question 19. Show that $\Pi(X, Y)$ is never empty. Show also that the infimum in (3) is always achieved. Is it unique?

Question 20. Let X, Y be two diagrams and ϕ be an optimal matching between them. We introduce $\gamma : t \mapsto \{(1-t)\, x + t\, \phi(x) \mid x \in X\}$. Show that γ is a geodesic between X and Y. Conclude.

7.4 Curvature of the space of persistence diagrams

An important step towards understanding the structure of a geodesic space is to study its curvature. The goal of this subsection is twofold: first, to prove that there is no upper bound on the curvature of the space of persistence diagrams; second, to prove that the curvature is everywhere lower-bounded by zero.

For the first objective, we use the following characterization of spaces with curvature bounded from above. If a geodesic space $(\mathcal{X}, d_\mathcal{X})$ has curvature upper-bounded by some $\kappa > 0$, then for any $x, y \in \mathcal{X}$ such that $d_\mathcal{X}(x, y) < \frac{1}{\kappa} \Rightarrow$ there is a *unique* geodesic between x and y.

Question 21. Using this characterization, exhibit a counterexample (or a family thereof) showing that there is no upper bound on the curvature of the space of persistence diagrams.

For the second objective, we use the following characterization: given a **geodesic** space $(\mathcal{X}, d_\mathcal{X})$, we say that $\mathcal{X}$ *is non-negatively curved* if for all $X, Y \in X$, for all geodesic $\gamma : [0,1] \to \mathcal{X}$ between X and Y, and any $Z \in \mathcal{X}$, we have:

$$\forall t \in [0,1], \quad d_\mathcal{X}(Z, \gamma(t))^2 \geqslant t\, d_\mathcal{X}(Z,Y)^2 + (1-t)\, d_\mathcal{X}(Z,X)^2 - t\,(1-t)\, d_\mathcal{X}(X,Y)^2 \tag{4}$$

When (4) appears to be an equality (for all X, Y, Z, γ, t), we say that $\mathcal{X}$ has *curvature zero.*

Question 22. Prove that $(\mathbb{R}^d, ||.||_2)$ has curvature zero.

We now consider three arbitrary diagrams, say X, Y, Z. Let ϕ_X^Y be an optimal matching between X and Y, and let γ be the corresponding geodesic as per Question 20. For $t \in [0,1]$, we introduce $\psi_t : X \to \mathbb{R}^2_> \cup \{\Delta\}$ defined by $\psi_t(x) = (1-t)\, x + t\, \phi_X^Y(x)$, and we also consider $\phi_Z^t : Z \to \gamma(t)$ an optimal matching between Z and $\gamma(t)$. We finally introduce $\phi_Z^X = (\psi_t)^{-1} \circ \phi_Z^t$ and $\phi_Z^Y = \phi_X^Y \circ \phi_Z^X$.

Question 23. Prove that, for any $x \in X$, if there is some $t \in (0,1)$ such that $\psi_t(x) = \Delta$, then $x = \Delta$ and $\phi_X^Y(x) = \Delta$.

Question 24. Prove the following results:

$$d_2(Z, \gamma(t))^2 = \sum_{z \in Z} ||z - [(1-t)\phi_Z^X(z) + t\phi_Z^Y(z)]||^2 \tag{5}$$

$$d_2(Z, Y)^2 \leqslant \sum_{z \in Z} ||z - \phi_Z^Y(z)||^2 \tag{6}$$

$$d_2(Z, X)^2 \leqslant \sum_{z \in Z} ||z - \phi_Z^X(z)||^2 \tag{7}$$

$$d_2(X, Y)^2 = \sum_{z \in Z} ||\phi_Z^X(z) - \phi_Z^Y(z)||^2 \tag{8}$$

Question 25. Combine Eqs (4-8) to conclude that the space of persistence diagrams is non-negatively curved.