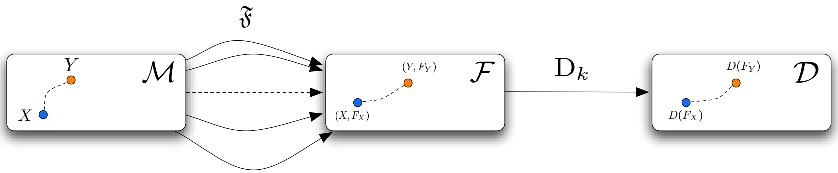
Curvature sets over Persistence Diagrams

Facundo Mémoli
Math and CSE Departments, OSU.
memoli@math.osu.edu

2014

Table of Contents

Persistent Homology pipeline



- $\mathcal{M}$: all finite metric spaces (X, d_X). Metrized with $d_{\mathcal{GH}}$.
- $\mathcal{F}$: all finite filtered spaces (X, F_X), $F_X : \mathrm{pow}(X) \to \mathbb{R}$ is a *filtration*.
- $\mathcal{D}$: all persistence diagrams. Metrized with bottlenck distance $d_{\mathcal{D}}$.
- $\mathfrak{F} : \mathcal{M} \to \mathcal{F}$ is a *filtration functor*. Think Rips, $\mathfrak{F}^R$.
- D_k: rank k Persistence Homology 'functor'.

Write

$$\mathrm{D}_k^R := \mathrm{D}_k \circ \mathfrak{F}^R : \mathcal{M} \to \mathcal{D}.$$

Motivation

In many applications one has huge datasets, say (X, d_X) with $\#X \simeq 10^6$. Computing $\mathrm{D}_k^{Rips}(X)$ is not feasible.

For instance, `jPlex` will struggle with $\#X = 500$.

An idea that has been implicitly/explicitly used is *bootstrapping/resampling*: Fix n (say 100) and compute

$$\mathrm{D}_k^R\big(\Psi_X^{(n)}(x_1, x_2, \ldots, x_n)\big)$$

for *many many choices* (say $N = 20,000$) of $x_1, x_2, \ldots, x_n$. Then compute *barycenter* of all those [MMH11].

This is *much cheaper* than attempting the computation of $\mathrm{D}_k^R(X)$ at once. For instance, we were forced to use this in a data intensive neuroscience application [SMI$^+$08].

These considerations lead to studying the object:

$$\mathbf{K}_{n,k}^R(X) := \big\{\mathrm{D}_k^R(M),\ M \in \mathbf{K}_n(X)\big\} \subset \mathcal{D}.$$

(Curvature sets over persistence diagrams)

Curvature sets

Let (X, d_X) compact metric space be given (the underlying dataset). Sample n-tuples from X and give them restriction metric:

$$\begin{aligned}\Psi_X^{(n)} : X \times \cdots \times X &\to \mathbb{R}_+^{n\times n}\\ (x_1,\ldots,x_n) &\mapsto ((d_X(x_i,x_j)))_{i,j=1}^n.\end{aligned}$$

Define the n-th **curvature set** of X as

$$\mathbf{K}_n(X) := \left\{\Psi_X^{(n)}(x_1,\ldots,x_n),\ (x_1,\ldots,x_n)\in X^{\times n}\right\} \subset \mathbb{R}_+^{n\times n}.$$

Think of this as the *n-point configuration space of X*. Contains all the building blocks of geometric simplicial complexes one can build from X.

Curvature sets and a distance on $\mathcal{M}$

Note that no matter what X, $\mathbf{K}_n(X)$ always lives as a subset of $\mathbb{R}_+^{n\times n}$.

It is therefore suggestive to try to compare X and Y by means of their respective curvature sets:

$$\widehat{d}_{\mathcal{GH}}(X,Y) := \frac{1}{2}\sup_{n\in\mathbb{N}} d_{\mathcal{H}}\big(\mathbf{K}_n(X),\mathbf{K}_n(Y)\big).$$

Is this a good definition (modified GH distance)?

Let's see some more details about curvature sets.

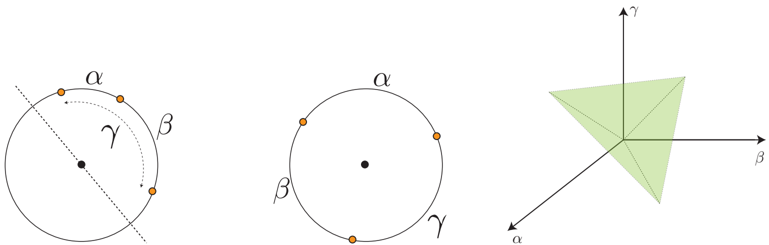
For example,

$$\mathbf{K}_2(X) = \left\{ \begin{pmatrix} 0 & d_X(x,x') \\ d_X(x,x') & 0 \end{pmatrix},\ x,x'\in X \right\}.$$

Curvature sets

$$\mathbf{K}_3(X) = \left\{ \begin{pmatrix} 0 & \alpha & \beta \\ \alpha & 0 & \gamma \\ \beta & \gamma & 0 \end{pmatrix}, \begin{array}{l} \alpha = d_X(x_1, x_2), \\ \beta = d_X(x_2, x_3), \\ \gamma = d_X(x_3, x_1), \end{array} x_1, x_2, x_3 \in X \right\}.$$

Example 2.1 **($\mathbf{K}_3(\mathbb{S}^1)$).** Consider $\mathbb{S}^1$ with angular metric. There are two possibilities for three points on $\mathbb{S}^1$. On the left figure: $\gamma = \alpha + \beta$, and on the right figure: $\alpha + \beta + \gamma = 2\pi$.



Total is 4 cases each given by a linear relation $\Rightarrow$ surface of tetrahedron with vertices $(0,0,0)$ and $\pi(1,1,0)$, $\pi(1,0,1)$, $\pi(0,1,1)$.

Curvature sets are functorial

$\mathbf{K}_n : \mathcal{M} \longrightarrow \mathrm{Borel}(\mathbb{R}_+^{n\times n})$ is functorial:

$$\mathbf{K}_n(X) \subseteq \mathbf{K}_n(Y) \text{ whenever } X \hookrightarrow Y \text{ isometrically.}$$

Consequence (all scaled down versions of $\mathbb{S}^1$ can be iso-embedded into $\mathbb{S}^2$):

$$\mathbf{K}_3(\mathbb{S}^2) = \mathrm{ConvexHull}(\mathbf{K}_3(\mathbb{S}^1)) \;\Rightarrow\; \text{full tetrahedron.}$$

Then,

$$\widehat{d}_{\mathcal{GH}}\left(\mathbb{S}^1, \mathbb{S}^2\right) \geqslant \frac{1}{2} d_{\mathcal{H}}\big(\mathbf{K}_3(\mathbb{S}^1), \mathbf{K}_3(\mathbb{S}^2)\big) = \frac{1}{2}\|\frac{\pi}{2}(1,1,1) - \frac{2\pi}{3}(1,1,1)\|_\infty = \frac{\pi}{12}.$$

Modified Gromov-Hausdorff distance

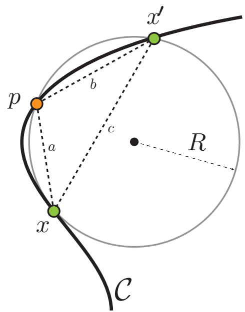
Theorem 2.1 ([Mém12]). *The following properties hold:*

- $\widehat{d}_{\mathcal{GH}}$ *is a legitimate distance on* $\mathcal{M}$ *modulo isometries.*
- $\widehat{d}_{\mathcal{GH}}$ *is topologically equivalent to* $d_{\mathcal{GH}}$ *on GH-precompact families.*
- $\widehat{d}_{\mathcal{GH}} \leqslant d_{\mathcal{GH}}$ *(and equality fails sometimes).*

Last property means that by looking at curvature sets we've proved that $d_{\mathcal{GH}}(\mathbb{S}^1, \mathbb{S}^2) \geqslant \frac{\pi}{12}$. It is typically difficult to find good lower bounds for GH between very symmetric spaces.

Why the name "curvature sets"?

Consider smooth plane curve $\mathcal{C}$:



Then, $\begin{pmatrix} 0 & a & b \\ a & 0 & c \\ b & c & 0 \end{pmatrix} \in \mathbf{K}_3(\mathcal{C})$, so from Heron's formula we can compute

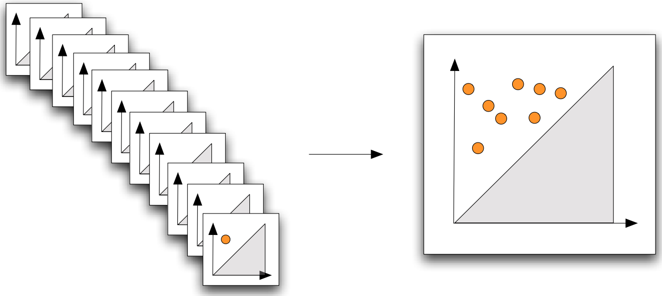
$$R^{-1} = \frac{4\,S(a,b,c)}{a\,b\,c}.$$

By a Taylor expansion $R^{-1} \simeq \kappa(p)$, the curvature at p, as x, x' and p coalesce.

Example: $\mathbf{K}^R_{4,1}(X)$

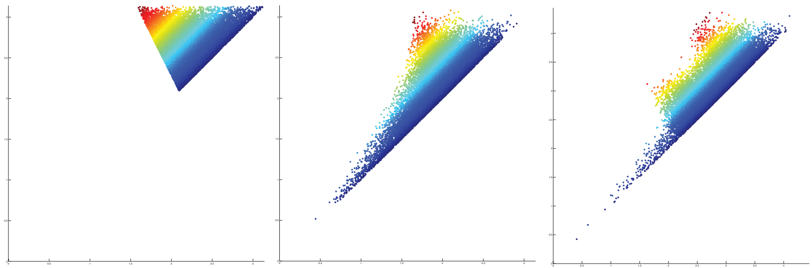
$$\mathbf{K}^R_{4,1}(X) := \{\mathrm{D}^R_1(M),\ M \in \mathbf{K}_4(X)\}.$$

Rips $\mathrm{D}_1(\cdot)$ of a space with 4 points can have *at most one point*. So I can take all $D \in \mathbf{K}^R_{4,1}(X)$ and plot them all in the same axis!



Example – cont'd

Consider $n = 4$ and $k = 1$ for $\mathbb{S}^1$, $\mathbb{S}^2$ and $\mathbb{T}^2$ (angular metric for all, and for Torus ℓ_2 mix). Computational example: sampled about 5×10^5 4-tuples out of each geometry. Color is *persistence* of each point.



Functoriality: $X \hookrightarrow Y$ isometrically implies $\mathbf{K}^R_{n,k}(X) \subseteq \mathbf{K}^R_{n,k}(Y)$.

In this case, $\mathbb{S}^1 \hookrightarrow \mathbb{S}^2$ and $\mathbb{S}^1 \hookrightarrow \mathbb{T}^2$. One can find the sub-diagram corresponding to $\mathbb{S}^1$ as a subset of the other two.

Curvature sets over PDs are well behaved

Recall that for *any* metric space X, $\mathbf{K}^R_{n,k}(X) \subset \mathcal{D}$. Thus, since D is a metric space (with the bottlenck distance), we can induce a Hausdorff distance on its subsets, $d^{\mathcal{D}}_{\mathcal{H}}$.

Theorem 3.1 (M-13). *For all X, Y compact metric spaces, and $k, n \in \mathbb{N}$, $n \geqslant 1$ it holds that:*

$$d^{\mathcal{D}}_{\mathcal{H}}\big(\mathbf{K}^R_{n,k}(X), \mathbf{K}^R_{n,k}(Y)\big) \leqslant 2\,\widehat{d}_{\mathcal{GH}}\,(X, Y)\,.$$

Remarks

- $\mathbf{K}^R_{n,k}(X)$ are stable *invariants* of a metric space/dataset.
- They are potentially as interesting as $\mathrm{D}^R_k(X)$.
- In many applications one is trying to discriminate datasets. Computing the lower bound to GH distance given by the theorem is very cheap. This is in sharp contrast with invoking standard stability of D^R_k to deduce a lower bound for GH distance.
- CSoPDs are combinatorial/set theoretic objects. They do not carry information that would allow me to answer the question: 'what is the most likely $D \in \mathbf{K}^R_{n,k}(X)$'?
- For that it is more natural to explore the idea of representing data as metric measure spaces $\Rightarrow$ induce *probability measures over* $\mathbf{K}^R_{n,k}(X)$.
- Understand probability measures over $\mathcal{D}$ $\Rightarrow$ model datasets as metric measure spaces.

Gromov-Wasserstein distance

Let $\mathcal{M}^w$ denote all compact mm-spaces: (X, d_X, μ_X) where μ_X is a Borel probability measures on X.

Isomorphism: $X \stackrel{\text{iso}}{=} Y$ if exists $\psi : X \to Y$ isometry such that $\psi_\# \mu_X = \mu_Y$.

Given (X, d_X, μ_X) and (Y, d_Y, μ_Y) in $\mathcal{M}^w$ let $\mu \in \mathcal{P}(X \times Y)$ be such that $(\pi_1)_\# \mu = \mu_X$ and $(\pi_2)_\# \mu = \mu_Y$. Any such μ is called a *coupling*. Define for each $p \geqslant 1$

$$d_{\mathcal{GW},p}(X, Y) := \frac{1}{2} \inf_{\mu} \| d_X - d_Y \|_{L^p(\mu \otimes \mu)}.$$

Theorem 4.1 ([Mém11]). *$d_{\mathcal{GW},p}$ is a legitimate distance on $\mathcal{M}^w \backslash \text{iso}$.*

Example 4.1. The one-point metric measure space is $(\{*\}, ((0)), \delta_*)$. Then,

$$d_{\mathcal{GW},p}(X, *) = \frac{1}{2} \left(\iint (d_X(x, x'))^p \mu_X(dx) \mu_X(dx') \right)^{1/p} =: \frac{1}{2} \mathbf{diam}_p (X).$$

This induces notion of convergence of a sequence $\{Z_n\} \subset \mathcal{M}^w$ to a point. *Concentration of measure*:

$$Z_n \xrightarrow{n} * \Leftrightarrow \mathbf{diam}_p (Z_n) \to 0 \text{ as } n \uparrow \infty.$$

Properties of $(\mathcal{M}^w, d_{\mathcal{GW},p})$

The space has been studied in [Mém11] and then by Sturm who in 2013 proved that

- $(\mathcal{M}^w, d_{\mathcal{GW},p})$ is a geodesic space, and
- for $p = 2$ it is an Alexandrov space of curvature $\geqslant 0$.

Probability measures on $\mathbf{K}^R_{n,k}(X)$

Pick any $(X, d_X, \mu_X) \in \mathcal{M}^w$.

Consider the map $\mathrm{D}^R_k \circ \Psi^{(n)}_X : \underbrace{X \times \cdots \times X}_{n \text{ times}} \longrightarrow \mathcal{D}$.

Then, consider product measure $\mu_X^{\otimes n}$ and push it forward via the map above and obtain a probability measure:

$$U_X^{(n,k)} := \left(\mathrm{D}^R_k \circ \Psi^{(n)}_X\right)_{\#} \mu_X^{\otimes n} \;\in\; \mathcal{P}(D).$$

Clearly, $\mathbf{K}^R_{n,k}(X) = \mathrm{supp}\left[U_X^{(n,k)}\right]$.

Now, given two metric measure spaces $X \mapsto U_X^{(n,k)}$ and $Y \mapsto U_Y^{(n,k)}$ are both elements of $\mathcal{P}(\mathcal{D})$.

The bottleneck distance on $\mathcal{D}$ induces a Wasserstein distance on $\mathcal{P}(\mathcal{D})$, $d^{\mathcal{D}}_{\mathcal{W},p}$.

Theorem 4.2 (M-13).

$$d^{\mathcal{D}}_{\mathcal{W},p}(U_X^{(n,k)}, U_Y^{(n,k)}) \leqslant c(n,k,p) \cdot d_{\mathcal{GW},p}(X,Y).$$

Cf. with Gromov-Prokhorov results by Blumberg et. al.

Concentration of measure

For each $n \in \mathbb{N}$ one can view $\mathbf{K}^R_{n,k}(X)$ as an mm-space, then obtain sequence

$$\left\{\left(\mathbf{K}^R_{n,k}(X), d_{\mathcal{D}}, U_X^{(n,k)}\right)\right\}_{n\in\mathbb{N}} \subset \mathcal{M}^w.$$

Does this sequence concentrate?

We can study whether $\lim_n \mathbf{diam}_p\left(\mathbf{K}^R_{n,k}(X)\right) = 0.$

As a consequence of mm-space covering theorem of [CM10] we obtain:

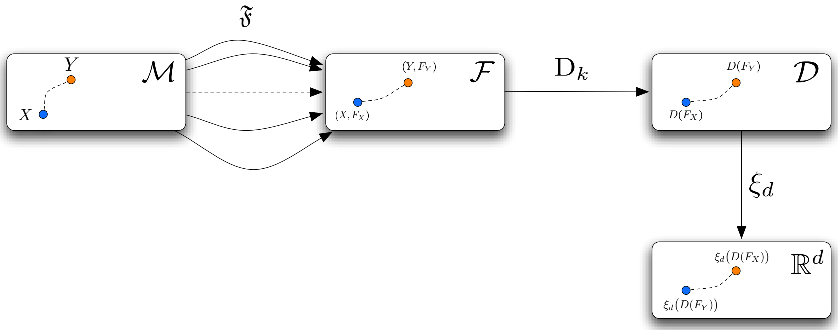
Theorem 4.3 (M-13).

$$d_{\mathcal{GW},p}(\mathbf{K}^R_{n,k}(X), *) \xrightarrow{n} 0$$

(Interpretation: convergence to barycenter.)

Cf. Related results using Gromov-Prokhorov distance by Blumberg et. al.

Coordinates on $\mathcal{D}$



$\xi_d : \mathcal{D} \to \mathbb{R}^d$ is a coordinatization map such that each coordinate is 1-Lipschitz.

Corollary 4.1. *For mm-spaces X and Y, and a 1-Lipschitz map $\xi : \mathcal{D} \to \mathbb{R}$:*

$$\int_0^\infty \left| P_{X,\xi}^{(n,k)}(t) - P_{Y,\xi}^{(n,k)}(t) \right| dt \leqslant c(n,k,p) \cdot d_{\mathcal{GW},p}(X,Y),$$

where $P_{X,\xi}^{(n,k)}(t) := U_X^{(n,k)}\Big(\big\{D \text{ s.t. } \xi(D) \leqslant t,\ D \in \mathbf{K}_{n,k}^R(X)\big\}\Big)$ *etc.*

Discussion

- Experimentation: discrimination of shapes using $\mathbf{K}^R_{n,k}(X)$.
- Precise (asymptotic) characterization of $U_X^{(n,k)}$ for $\mathbb{S}^1$, $\mathbb{S}^2$, $\mathbb{T}^2$?
- Structural properties of $\mathbf{K}^R_{n,k}(X)$, or $U_X^{(n,k)}$?
- preprint should be posted on Arxiv soon.

Bibliography

References

[CM10] Gunnar Carlsson and Facundo Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11:1425–1470, 2010.

[Mém11] Facundo Mémoli. Gromov-Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, pages 1–71, 2011. 10.1007/s10208-011-9093-5.

[Mém12] Facundo Mémoli. Some properties of Gromov-Hausdorff distances. *Discrete & Computational Geometry*, pages 1–25, 2012. 10.1007/s00454-012-9406-8.

[MMH11] Yuriy Mileyko, Sayan Mukherjee, and John Harer. Probability measures on the space of persistence diagrams. *Inverse Problems*, 27(12):124007, 2011.

[SMI+08] Gurjeet Singh, Facundo Memoli, Tigran Ishkhanov, Guillermo Sapiro, Gunnar Carlsson, and Dario L. Ringach. Topological analysis of population activity in visual cortex. *Journal of Vision*, 8(8), 2008.