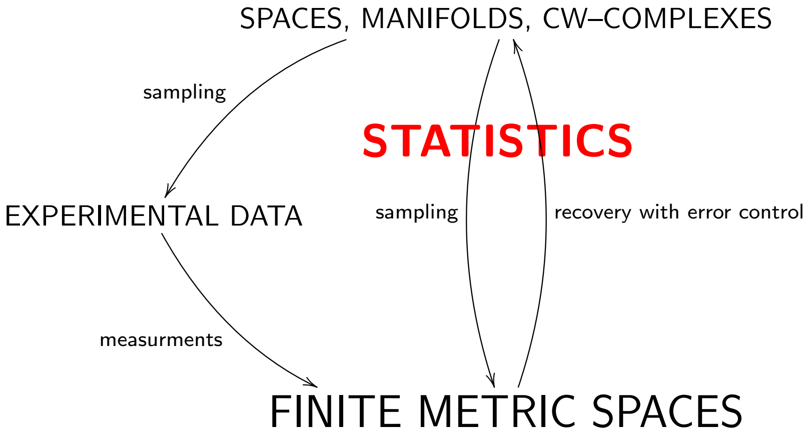
Multi-persistence, Saas 2016

TDA group, KTH Stockholm

Anders Lundman, Ryan Ramanujam
Martina Scolamiero, Sebastian Öberg, Wojciech Chachólski, ...

Applied Topology



Gromov-Hausdorff distance. It requires $n!$ computations.

Data system: a set U with measurements $\{U \xrightarrow{m_i} (X_i, d_i)\}_{i=1}^n$.

Statistical methods require small errors and precise measurements.

More and more data is:

- acquired by processes with unavoidable errors,
- acquired by measurements done using equipment with different technology and incomparable protocols,
- heterogenous.

Need methods that:

- reduce the dependence on the metric;
- extract properties preserved by for example rescaling.

Need homotopy theory of data systems

$$U \xmapsto{\text{Invariants}} I(U)$$

- whose values are "easily" comparable for different U's
- visualizable
- CONTINUOUS
- feasible to calculate
- hopefully the outcome is suitable for statistical analysis

An invariant is a simplification process.

Topological Data analysis is the study of such invariants obtained from homological calculations.

Starting point:

Data system: a set U with measurements $\{U \xrightarrow{m_i} (X_i, d_i)\}_{i=1}^n$.

Recovery with error information:

For $(r_1, \ldots, r_n) \in \mathbf{Q}^n_{\geq 0}$, define a simplicial complex $\mathbf{U}(r_1, \ldots, r_n)$:

- U is the set its 0-simplices
- $(x_0, \ldots, x_r) \in U^{r+1}$ is an r-simplex if there is a point y in U such that $d_i(x_i, y) \leq r_i$.

This leads to a persistence space $\mathbf{U} \colon \mathbf{Q}^n_{\geq 0} \to$ Spaces.

Homological Invariants:

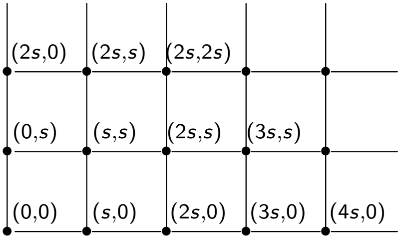
Apply homology to obtain a persistence module:

$$H_i(\mathbf{U}) \colon \mathbf{Q}^n_{\geq 0} \to \mathrm{Vect}_K$$

Obtained persistence modules $H\colon \mathbf{Q}^n_{\geq 0} \to \mathrm{Vect}_K$ are quite special:

- they are compact objects in $\mathrm{Fun}(\mathbf{Q}^n_{\geq 0}, \mathrm{Vect}_K)$;
- and tame: there is $G\colon \mathbf{N}^n \to \mathrm{Vect}_K$ and $s \in \mathbf{Q}_{>0}$, such that H is isomorphic to the left Kan extension of G along the scaled lattice $s\mathbf{N}^n \subset \mathbf{Q}^n_{\geq 0}$.

For example, for $n = 2$, H is tame if, for some s, it is constant on the half open squares $[as, (a+1)s) \times [bs, (b+1)s)$ for a, b in $\mathbf{N}$:



Example of tame and compact persistence modules

Free on one generator. Let $v \in \mathbf{Q}^n_{\geq 0}$. Set $[v, \infty)\colon \mathbf{Q}^n_{\geq 0} \to \mathrm{Vect}_K$

$$[v,\infty)(w) = \begin{cases} K & \text{if } v \leq w \\ 0 & \text{if } v \not\leq w \end{cases}$$

non-zero maps
between non-zero values

Free. Let $\beta\colon \mathbf{Q}^n_{\geq 0} \to \mathbf{N}$ be function. It is called the Betti diagram of the free functor:

$$\bigoplus_{v\in \mathbf{Q}^n_{\geq 0}} [v,\infty)^{\beta v}$$

β has finite support if and only if this functor is compact.

Bar. For $v \leq w \in \mathbf{Q}^n_{\geq 0}$, the cokernel of the unique inclusion $[w, \infty) \subset [v.\infty)$ is denoted denoted by $[v, w)$.

The category of tame functors $\mathrm{Tame}(\mathbf{Q}^n_{\geq 0}, \mathrm{Vect}_K)$ is like the category of multi graded modules over $K[X_1, \ldots, X_n]$:

- It is Abelian with enough projectives,
- where all projectives are free.
- It is of dimension n: all tame persistence modules have a projective (free) resolution of length at most n.
- For $n = 1$, any tame and compact persistence module is isomorphic to a direct sum of bars $\oplus_{i=1}^{k}[v_i, w_i)$.

$$\text{Data Systems} \longrightarrow \text{Tame}(\mathbf{Q}^n_{\geq 0}, \text{Vect}_K)$$

$$U \longmapsto H_i\mathbf{U}$$

Looking after invariants:

- whose values are computationally comparable for different U's
- visualizable
- CONTINUOUS
- feasible to calculate

$$\text{Data Systems} \longrightarrow \text{Tame}(\mathbf{Q}^n_{\geq 0}, \text{Vect}_K)$$

$$U \longmapsto H_i\mathbf{U}$$

OK $n = 1$ whose values are computationally comparable for different U's

NO $n > 1$ whose values are computationally comparable for different U's

OK $n = 1$ visualizable

NO $n > 1$ visualizable

? CONTINUOUS

OK feasible to calculate

Need to simplify further

- Betti diagrams and Euler characteristics:

$$\mathrm{Tame}(\mathbf{Q}^n_{\geq 0}, \mathrm{Vect}_K) \longrightarrow \{\text{Functions } \mathbf{Q}^n_{\geq 0} \to \mathbf{N}\}$$

$$F \longmapsto \beta_i(F)$$

$$F \longmapsto \chi(F) = \Sigma^n_{i=0} \beta_i(F)$$

- Ranks of Betti diagrams:

$$\mathrm{Tame}(\mathbf{Q}^n_{\geq 0}, \mathrm{Vect}_K) \longrightarrow \mathbf{N}$$

$$F \longmapsto \mathrm{rank}_i F := \Sigma_{v \in \mathbf{Q}^n_{\geq 0}} \beta_i(F)(v)$$

- Minimal sets of gen (topological features of the data system)

$$F \longmapsto \text{a minimal set of generators for } F$$

Extract continuous invariants of tame persistence modules?

Definition. A Noise is a sequence $\{\mathcal{S}_\epsilon\}_{\epsilon \in \mathbf{Q}_{\geq 0}}$ of collections of tame persistence modules, called components. Elements in $\mathcal{S}_\epsilon$ are called ϵ-small. These collections are required to satisfy:

- $0 \in \mathcal{S}_\epsilon$ for any ϵ;
- $\mathcal{S}_\tau \subset \mathcal{S}_\epsilon$ if $\tau \leq \epsilon$;
- If $0 \to F \to H \to G \to 0$ is an exact sequence of tame persistence modules, then:
  - if $H \in \mathcal{S}_\epsilon$, then $F, G \in \mathcal{S}_\epsilon$;
  - if $F \in \mathcal{S}_\tau$ and $G \in \mathcal{S}_\epsilon$, then $H \in \mathcal{S}_{\tau+\epsilon}$.

Collection $\cup_{t>\epsilon}\mathcal{S}_t \subset \cup_{t\geq\epsilon}\mathcal{S}_t$ are Serre classes.

Standard Noise in the direction of a cone.

Let V be a subset $\mathbf{Q}^n_{\geq 0}$. Define:

$$\mathcal{V}_\epsilon = \left\{ F \text{ tame} \mid \text{for any } x \in F(u) \begin{array}{c} \text{there is } v \in V \text{ s.t. } v_i \leq \epsilon \\ x \in \mathrm{Ker}(F(u) \to F(u+v)) \end{array} \right\}$$

If V is a cone, then $\{\mathcal{V}_\epsilon\}_{\epsilon \in \mathbf{Q}_{\geq 0}}$ is a noise.

Domain noise. Let $\{X_\epsilon\}_{\epsilon \in \mathbf{Q}_{\geq 0}}$ be a sequence of subsets of $\mathbf{Q}^n_{\geq 0}$.

Define:

$$\mathcal{X}_\epsilon = \{F \text{ tame} \mid \text{if } F(u) \neq 0, \text{ then } u \in X_\epsilon\}$$

If $X_\tau \subset X_\epsilon$ for any $\tau \leq \epsilon$, then $\{\mathcal{X}_\epsilon\}_{\epsilon \in \mathbf{Q}_{\geq 0}}$ is a noise.

Dimension noise. Let $\{k_\epsilon\}_{\epsilon\in\mathbf{Q}_{\geq 0}}$ be a sequence of natural numbers. Define:

$$\mathcal{N}_\epsilon = \left\{F \text{ tame} \mid \dim F(u) \leq k_\epsilon \text{ for any } u \in \mathbf{Q}^n_{\geq 0}\right\}$$

If $k_0 = 0$ and $k_\tau + k_\epsilon \leq k_{\tau+\epsilon}$, then $\{\mathcal{N}_\epsilon\}_{\epsilon\in\mathbf{Q}_{\geq 0}}$ is a noise.

Intersection noise. If $\{\mathcal{S}_\epsilon\}_{\epsilon\in\mathbf{Q}_{\geq 0}}$ and $\{\mathcal{T}_\epsilon\}_{\epsilon\in\mathbf{Q}_{\geq 0}}$ are noises, then so is their intersection: $\{\mathcal{S}_\epsilon \cap \mathcal{T}_\epsilon\}_{\epsilon\in\mathbf{Q}_{\geq 0}}$.

Noise generated by a functor. Let F be a tame persistence module and $\alpha \in \mathbf{Q}_{\geq 0}$. Define $< F, \alpha >$ to be the intersection of all the noises $\{\mathcal{S}_\epsilon\}_{\epsilon\in\mathbf{Q}_{\geq 0}}$ that contain F in $\mathcal{S}_\alpha$.

For $r = 1$, up to rescale, there is only one noise whose components are closed under direct sums.

Topology and metric on tame persistence modules.

Let $\{\mathcal{S}_\epsilon\}_{\epsilon \in \mathbf{Q}_{\geq 0}}$ be a noise.
Two tame persistence modules F and G are α-close if there are natural transformations: $F \xleftarrow{f} H \xrightarrow{g} G$. such that:

$$\ker(f) \in \mathcal{S}_{\tau_1},\ \operatorname{coker}(f) \in \mathcal{S}_{\tau_2}, \ker(g) \in \mathcal{S}_{\tau_3},\ \ \ker(g) \in \mathcal{S}_{\tau_4}$$

$$\tau_1 + \tau_2 + \tau_3 + \tau_4 < \alpha$$

This defines a metric on tame persistence modules.
Discs:

$$B(F, \alpha) = \{G \text{ tame } | G \text{ is } \alpha\text{-close to } F\}$$

form a base for a topology on the collection of tame persistence modules.

This topology can be used to construct a continuous invariant of a persistence module.

Let F be a tame persistence module.

A bar code of F is a function $\text{bar}_i(F)\colon \mathbf{R}_{\geq 0} \to \mathbf{N}$ defined as:

$$\text{bar}_i(F)_\alpha := \min\{\text{rank}_i G \mid G \in B(F, \alpha)\}$$

A bar code is a non increasing function that measures how many features stay alive after a time α.

Theorem.

$$\begin{array}{ccc} \text{Tame}(\mathbf{Q}^n_{\geq 0}, \text{Vect}_K) & \longrightarrow & \{\text{Functions } \mathbf{R} \to \mathbf{N}\} \\ F & \longmapsto & \text{bar}_0(F) \end{array}$$

is a continuous function (in fact 1-Lipschitz).

$$\begin{array}{ccc} \text{Tame}(\mathbf{Q}^n_{\geq 0}, \text{Vect}_K) & \longrightarrow & \{\text{Functions } \mathbf{R} \to \mathbf{N}\} \\ F & \longmapsto & \text{bar}(F) \end{array}$$

OK values are computationally comparable for different F's

OK visualizable

OK CONTINUOUS

? feasible to calculate

- If $n = 1$ and the standard noise, bar(F) recovers the usual 1-persistence bar code.
- If $n > 1$, then calculating bar($-$) is an NP hard problem.

Data system:

- a set U with measurements $\{U \xrightarrow{m_i} (X_i, d_i)\}_{i=1}^{n}$
- a choice of a noise