Bremen - July, 18 2013

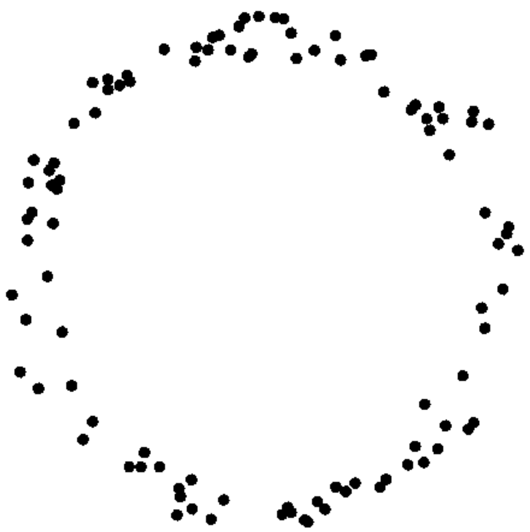
Applied and Computational Algebraic Topology

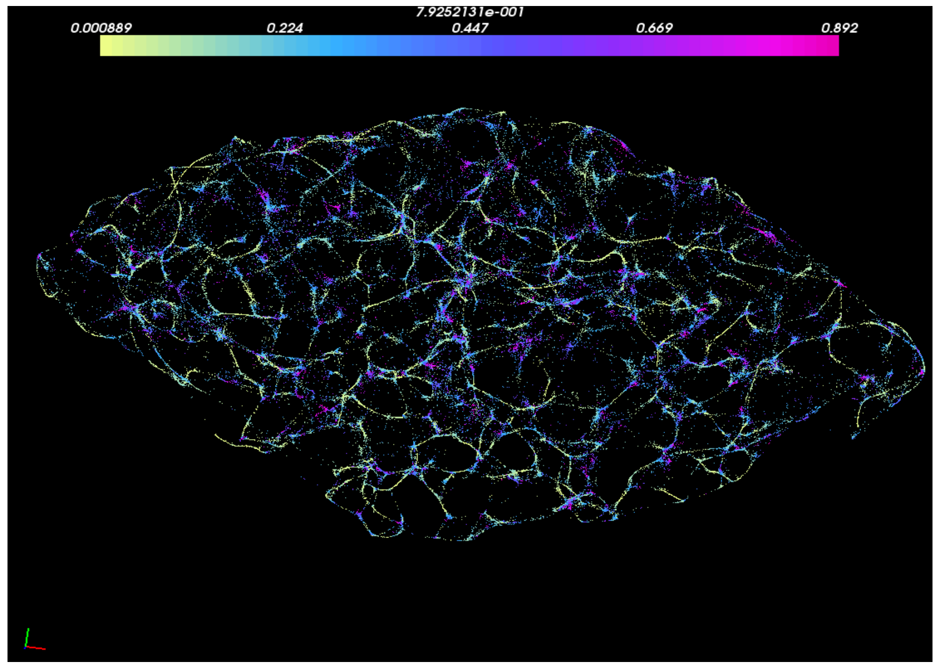
Optimal rates of convergence for persistence diagrams in Topological Data Analysis

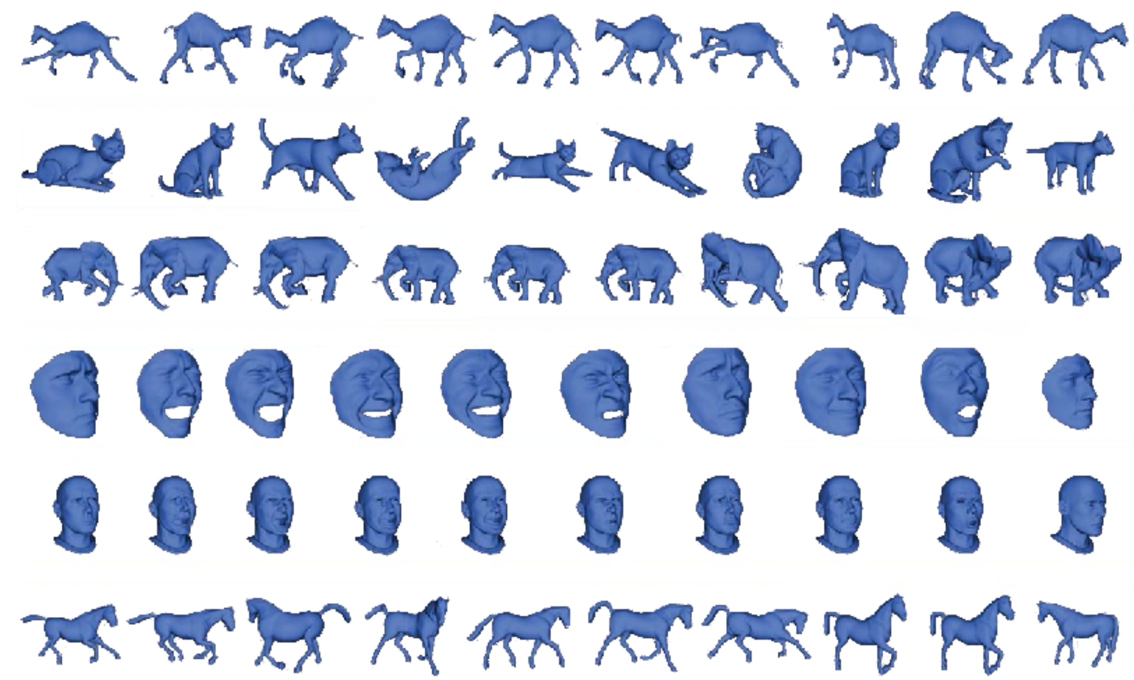
Frédéric Chazal

Joint work with M. Glisse, C. Labruère and B. Michel
(+ V. de Silva and S. Oudot)

Introduction

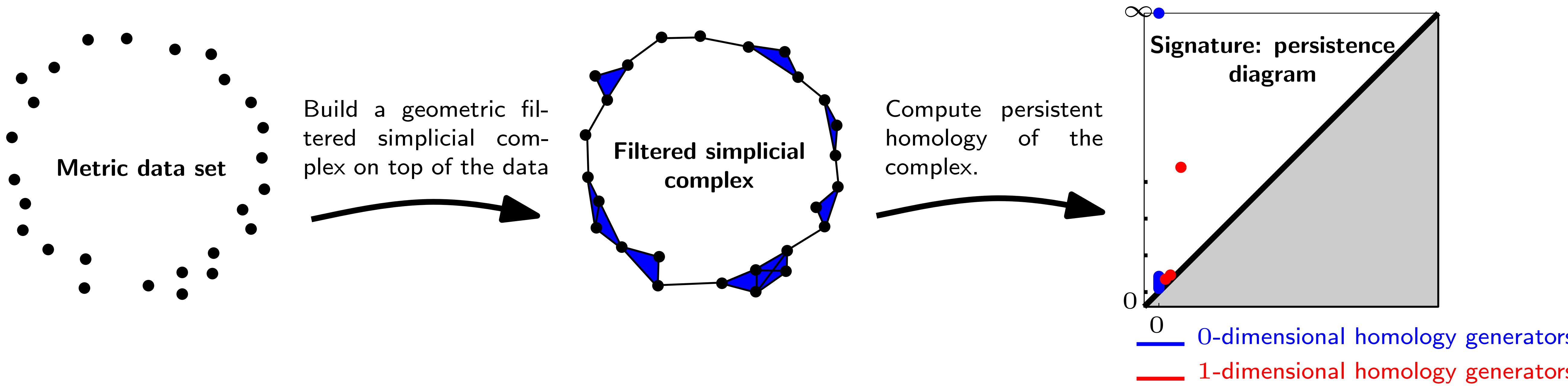






- Data often come as (sampling of) metric spaces or sets/spaces endowed with a similarity measure with possibly complex topological/geometric structure.
- TDA: infer relevant topological and geometric features of these spaces.

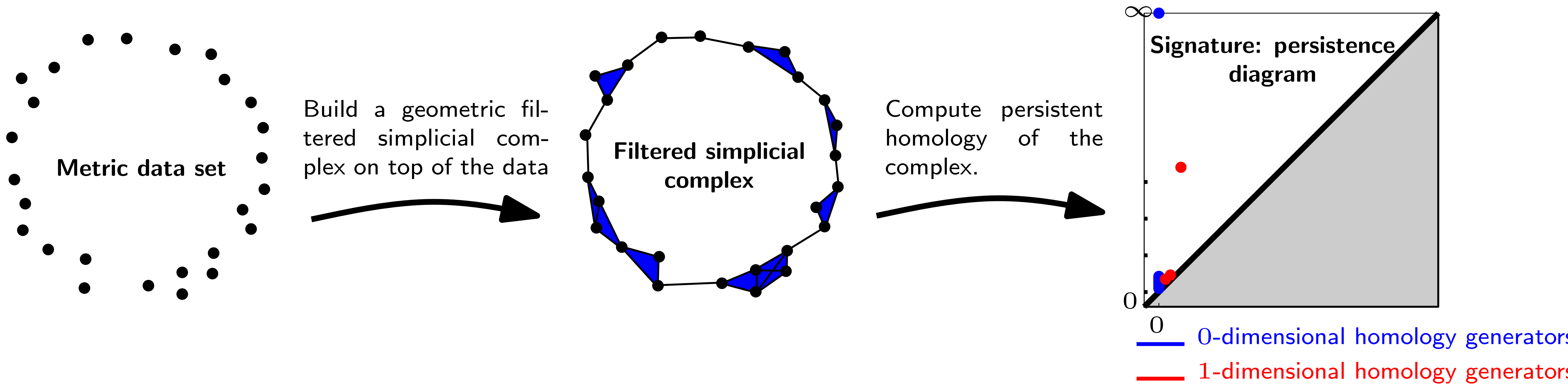
Topological signatures for data



A “classical” approach:

- Build a geometric filtered simplicial complex on top of $(\mathbb{X}, \rho_{\mathbb{X}})$ (ρ_X being a metric/similarity on $\mathbb{X}$).

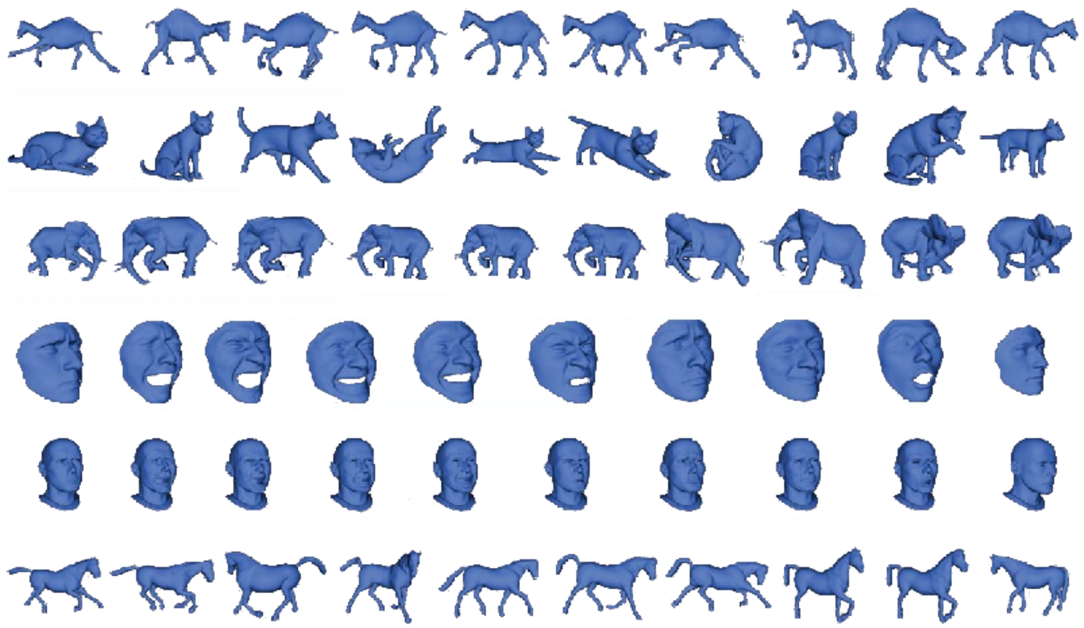
Topological signatures for data



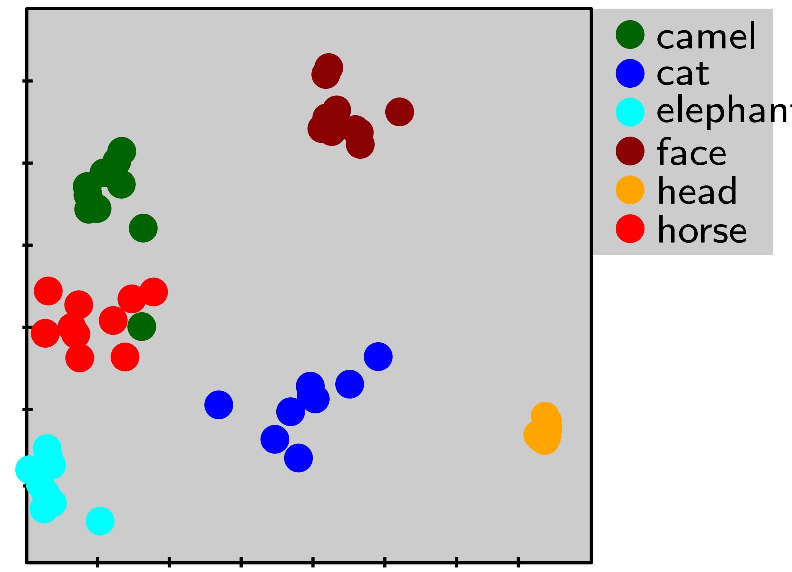
A "classical" approach:

- Build a geometric filtered simplicial complex on top of $(\mathbb{X}, \rho_{\mathbb{X}})$ (ρ_X being a metric/similarity on $\mathbb{X}$).
- Compute the persistent homology of the complex $\rightarrow$ persistence diagrams: multiscale topological signature.

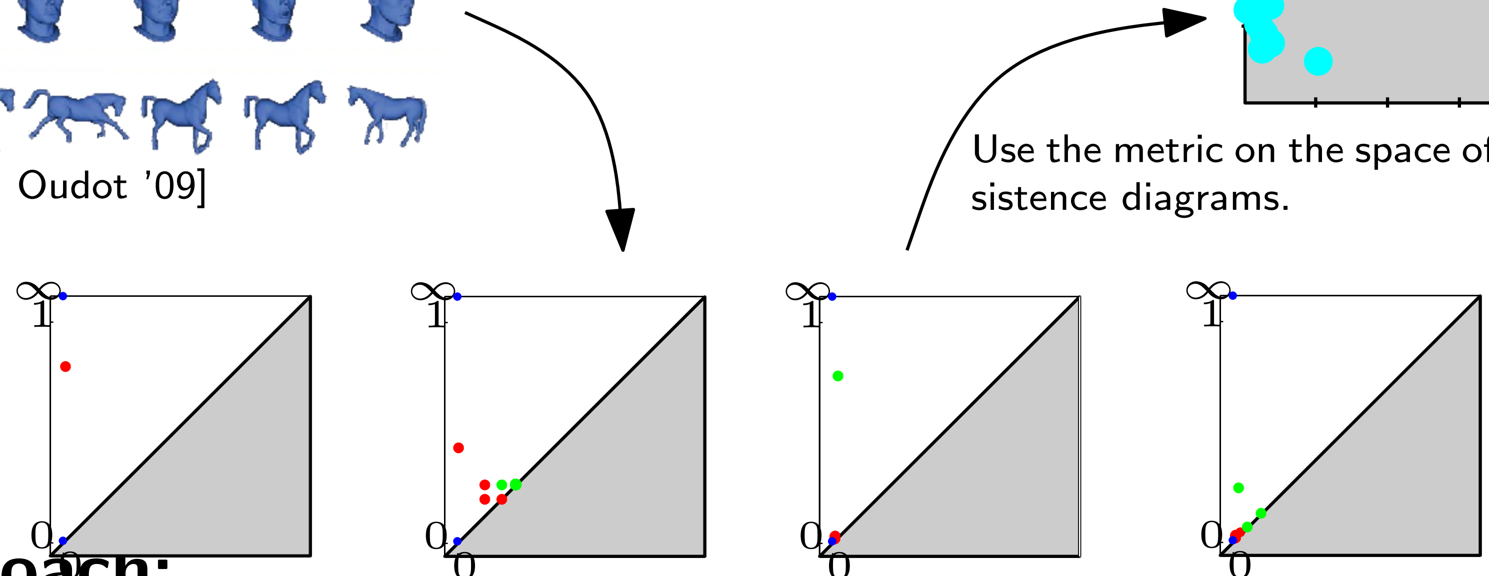
Topological signatures for data



[C., Cohen-Steiner, Guibas, Mémoli, Oudot '09]



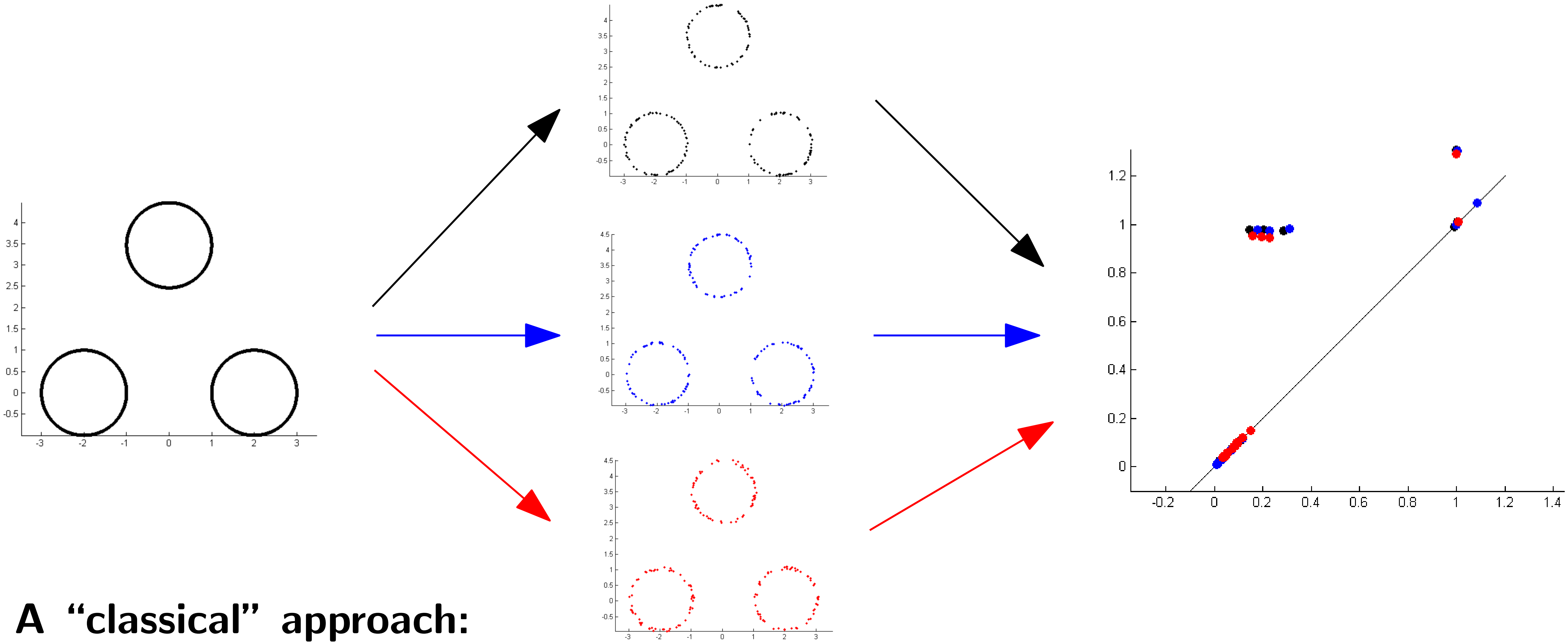
Use the metric on the space of persistence diagrams.



A "classical" approach:

- Build a geometric filtered simplicial complex on top of $(\mathbb{X}, \rho_{\mathbb{X}})$ (ρ_X being a metric/similarity on $\mathbb{X}$).
- Compute the persistent homology of the complex $\to$ persistence diagrams: multiscale topological signature.
- Compare the signatures of "close" data sets $\to$ robustness and stability results.

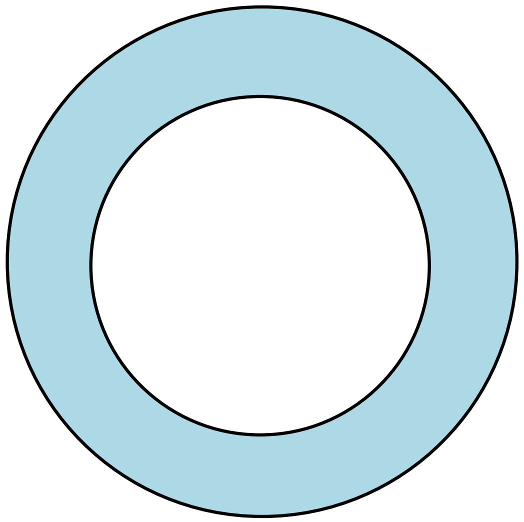
Topological signatures for data



A "classical" approach:

- Build a geometric filtered simplicial complex on top of $(\mathbb{X}, \rho_{\mathbb{X}})$ (ρ_X being a metric/similarity on $\mathbb{X}$).
- Compute the persistent homology of the complex $\rightarrow$ persistence diagrams: multiscale topological signature.
- Compare the signatures of "close" data sets $\rightarrow$ robustness and stability results.
- Statistical properties of signatures?

Topological signatures for data



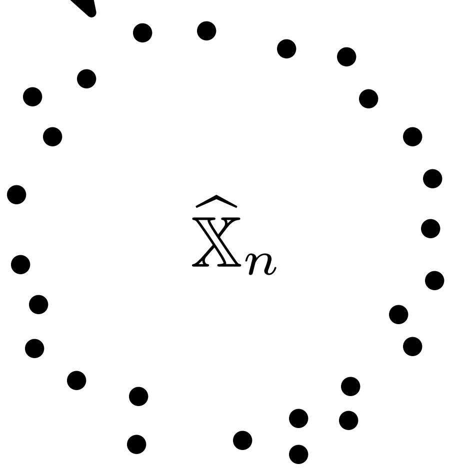
$(\mathbb{M}, \rho)$ metric space

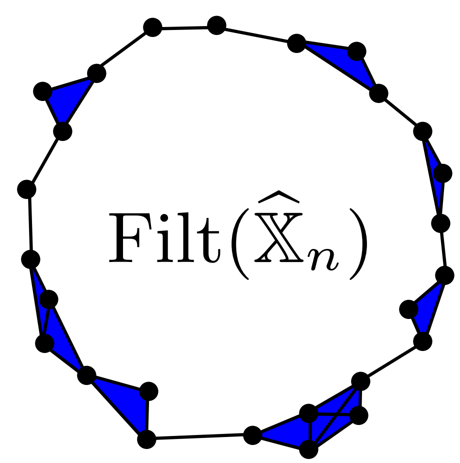
μ a probability measure with compact support $\mathbb{X}_\mu$.

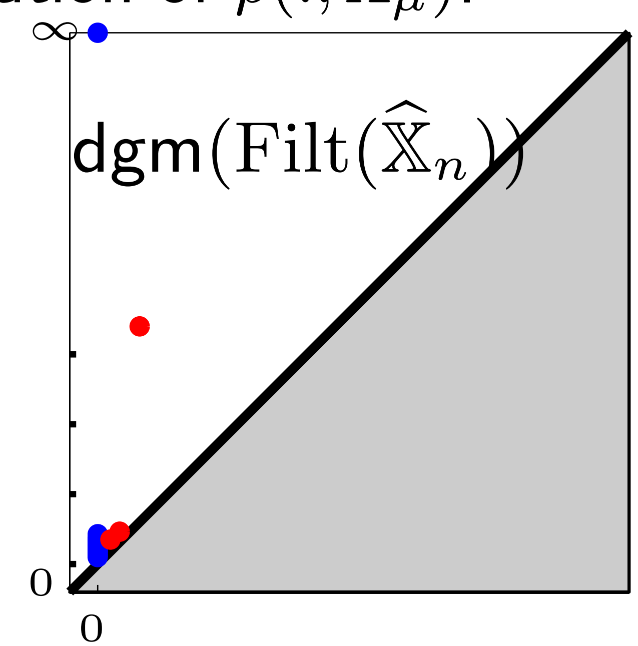
Examples:

- $\mathrm{Filt}(\widehat{\mathbb{X}}_n) = \mathrm{Rips}_\alpha(\widehat{\mathbb{X}}_n)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_n) = \check{\mathrm{C}}\mathrm{ech}_\alpha(\widehat{\mathbb{X}}_n)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_n) =$ sublevelset filtration of $\rho(., \mathbb{X}_\mu)$.

Sample n points according to μ.



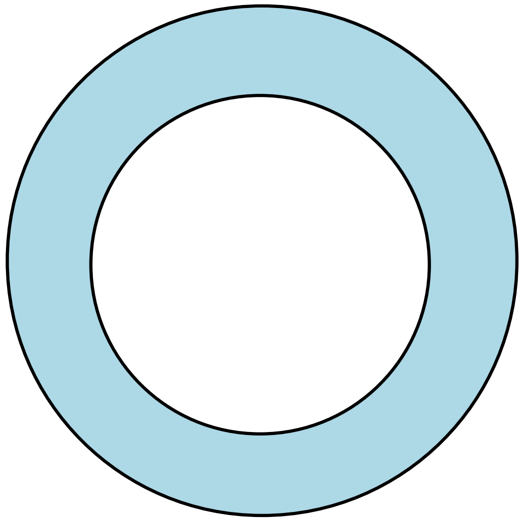




Questions:

- Statistical properties of dgm$(\mathrm{Filt}(\widehat{\mathbb{X}}_n))$? dgm$(\mathrm{Filt}(\widehat{\mathbb{X}}_n)) \to ?$ as $n \to +\infty$?

Topological signatures for data



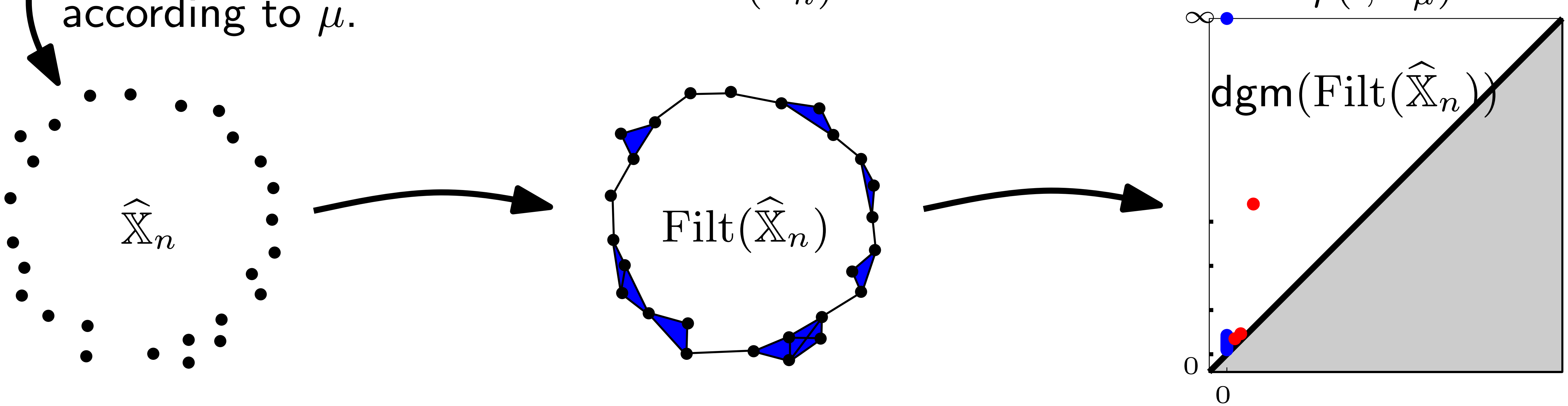
$(\mathbb{M}, \rho)$ metric space

μ a probability measure with compact support $\mathbb{X}_\mu$.

Examples:

- $\mathrm{Filt}(\widehat{\mathbb{X}}_n) = \mathrm{Rips}_\alpha(\widehat{\mathbb{X}}_n)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_n) = \check{\mathrm{C}}\mathrm{ech}_\alpha(\widehat{\mathbb{X}}_n)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_n) =$ sublevelset filtration of $\rho(., \mathbb{X}_\mu)$.

Sample n points according to μ.



Questions:

- Statistical properties of $\mathrm{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_n))$? $\mathrm{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_n)) \to ?$ as $n \to +\infty$?
- Is $\mathrm{dgm}(\mathrm{Filt}(\mathbb{X}_\mu))$ well-defined? (not obvious, even when $\mathbb{X}_\mu$ is a compact smooth submanifold of $\mathbb{R}^d$) Stability properties?

$\mathrm{Filt}(\mathbb{X}_\mu)$: persistence and stability of filtrations built on top of (pre)compact metric spaces

Filtered complexes

A filtered simplicial complex $\mathbb{S}$ built on top of a set $\mathbb{X}$ is a family $(\mathbb{S}_a \mid a \in \mathbf{R})$ of subcomplexes of some fixed simplicial complex $\overline{\mathbb{S}}$ with vertex set X s. t. $\mathbb{S}_a \subseteq \mathbb{S}_b$ for any $a \leq b$.

Examples: Let $(\mathbb{X}, \rho)$ be a metric space.

- The Vietoris-Rips and Čech complexes $\mathrm{Rips}(\mathbb{X})$ and $\check{\mathrm{C}}\mathrm{ech}(\mathbb{X})$ are the filtered complexes defined by: for $a \in \mathbf{R}$,

$$[x_0, x_1, \cdots, x_k] \in \mathrm{Rips}(\mathbb{X}, a) \Leftrightarrow \rho(x_i, x_j) \leq a, \quad \text{for all } i, j$$

$$[x_0, x_1, \ldots, x_k] \in \check{\mathrm{C}}\mathrm{ech}(\mathbb{X}, a) \quad \Leftrightarrow \quad \bigcap_{i=0}^{k} B(x_i, a) \neq \emptyset,$$

where $B(x, a) = \{x' \in \mathbb{X} : \rho(x, x') \leq a\}$.

Tame persistent modules

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Examples:

- Let $\mathbb{S}$ be a filtered simplicial complex. If $V_a = \mathrm{H}(\mathbb{S}_a)$ and $v_a^b : \mathrm{H}(\mathbb{S}_a) \to \mathrm{H}(\mathbb{S}_b)$ is the linear map induced by the inclusion $\mathbb{S}_a \hookrightarrow \mathbb{S}_b$ then $(\mathrm{H}(\mathbb{S}_a) \mid a \in \mathbf{R})$ is a persistence module.
- Given a metric space $(\mathbb{X}, \rho)$, $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ is a persistence module.
- Given a metric space $(\mathbb{X}, \rho)$, $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ is a persistence module.
- If $\mathbb{X} \subset (\mathbb{M}, \rho)$ and $d_{\mathbb{X}} = \rho(., \mathbb{X})$, then $(\mathrm{H}(d_{\mathbb{X}}^{-1}([0, a])) \mid a \in \mathbf{R})$ is a persistence module.

Tame persistent modules

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Definition: A persistence module $\mathbb{V}$ is q-tame if for any $a < b$, v_a^b has a finite rank.

Theorem [CCGGO'09-CdSGO'12]:
q-tame persistence modules have well-defined persistence diagrams.

Tame persistent modules

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Definition: A persistence module $\mathbb{V}$ is q-tame if for any $a < b$, v_a^b has a finite rank.

Theorem [CCGGO'09-CdSGO'12]:
q-tame persistence modules have well-defined persistence diagrams.

Theorem[CdSO'12]: Let $\mathbb{X}$ be a precompact metric space. Then $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ are q-tame.

As a consequence dgm$(\mathrm{H}(\mathrm{Rips}(\mathbb{X})))$ and dgm$(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X})))$ are well-defined!

Recall that a metric space $(\mathbb{X}, \rho)$ is precompact if for any $\epsilon > 0$ there exists a finite subset $F_\epsilon \subset \mathbb{X}$ such that $d_H(\mathbb{X}, F_\epsilon) < \epsilon$ (i.e. $\forall x \in X, \exists p \in F_\epsilon$ s.t. $\rho(x,p) < \epsilon$).

Tame persistent modules

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

A homomorphism of degree ϵ between two persistence modules $\mathbb{U}$ and $\mathbb{V}$ is a collection Φ of linear maps

$$(\phi_a : U_a \to V_{a+\epsilon} \mid a \in \mathbf{R})$$

such that $v_{a+\epsilon}^{b+\epsilon} \circ \phi_a = \phi_b \circ u_a^b$ for all $a \leq b$.

$$\begin{tikzcd}
U^a \arrow[rr] \arrow[dr] & & U^b \arrow[dr] & \\
& V^{a+\epsilon} \arrow[rr] \arrow[ur, phantom, "\circlearrowleft"] & & V^{b+\epsilon}
\end{tikzcd}$$

An ε-interleaving between $\mathbb{U}$ and $\mathbb{V}$ is specified by two homomorphisms of degree ϵ $\Phi : \mathbb{U} \to \mathbb{V}$ and $\Psi : \mathbb{V} \to \mathbb{U}$ s.t. $\Phi \circ \Psi$ and $\Psi \circ \Phi$ are the "shifts" of degree 2ϵ between $\mathbb{U}$ and $\mathbb{V}$.

$$\begin{tikzcd}
\cdots \arrow[rr] & & U^a \arrow[rr, "u_a^{a+2\epsilon}"] \arrow[dr, "\phi_a"'] & & U^{a+2\epsilon} \arrow[rr] \arrow[dr] & & \cdots & \\
& \cdots \arrow[ur] \arrow[rr] & & V^{a+\epsilon} \arrow[ur, "\psi_{a+\epsilon}"'] \arrow[rr, "v_{a+\epsilon}^{a+3\epsilon}"'] & & V^{a+3\epsilon} \arrow[rr] & & \cdots
\end{tikzcd}$$

Tame persistent modules

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Stability Theorem [CCGGO'09-CdSGO'12]:
If $\mathbb{U}$ and $\mathbb{V}$ are q-tame and ϵ-interleaved for some $\epsilon \geq 0$ then

$$d_B(\mathsf{dgm}(\mathbb{U}), \mathsf{dgm}(\mathbb{V})) \leq \epsilon$$

Tame persistent modules

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Stability Theorem [CCGGO'09-CdSGO'12]:
If $\mathbb{U}$ and $\mathbb{V}$ are q-tame and ϵ-interleaved for some $\epsilon \geq 0$ then

$$d_B(\mathsf{dgm}(\mathbb{U}), \mathsf{dgm}(\mathbb{V})) \leq \epsilon$$

Strategy: build filtered complexes on top of metric spaces that induce **q-tame** homology persistence modules and that turns out to be ϵ-**interleaved** when the considered spaces are $O(\epsilon)$-*close*.

Tame persistent modules

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

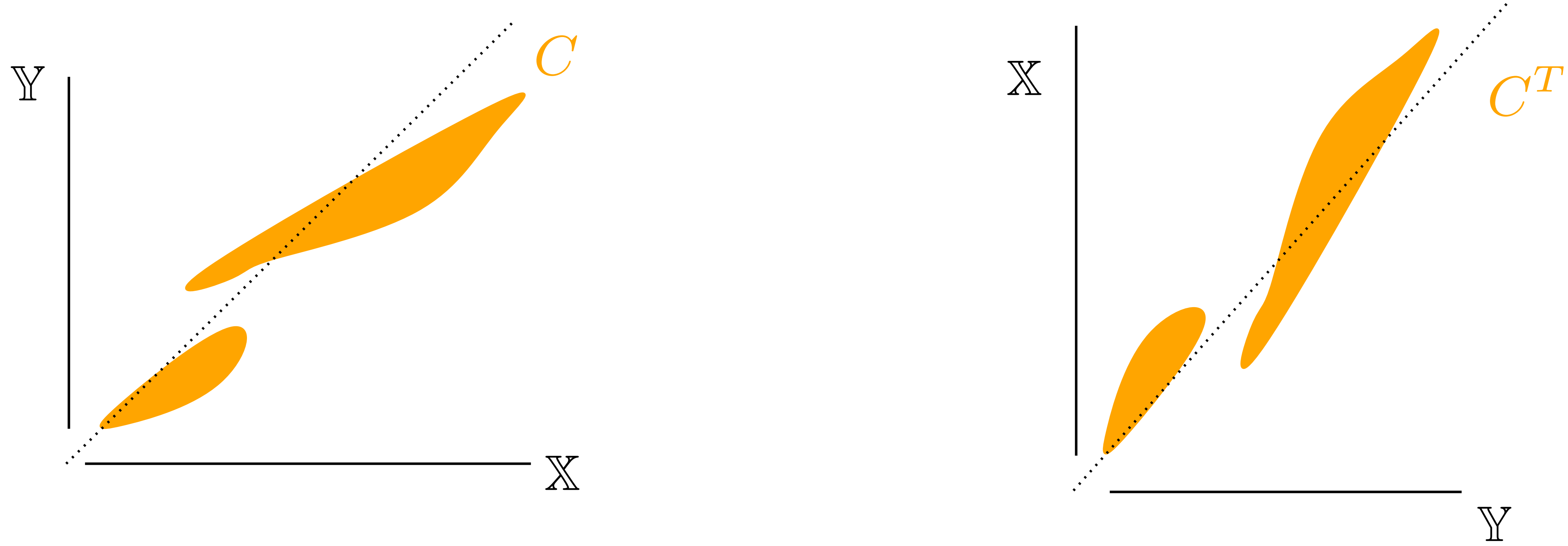
Stability Theorem [CCGGO'09-CdSGO'12]:
If $\mathbb{U}$ and $\mathbb{V}$ are q-tame and ϵ-interleaved for some $\epsilon \geq 0$ then

$$d_B(\text{dgm}(\mathbb{U}), \text{dgm}(\mathbb{V})) \leq \epsilon$$

Strategy: build filtered complexes on top of metric spaces that induce **q-tame** homology persistence modules and that turns out to be ϵ-**interleaved** when the considered spaces are $O(\epsilon)$-*close*.

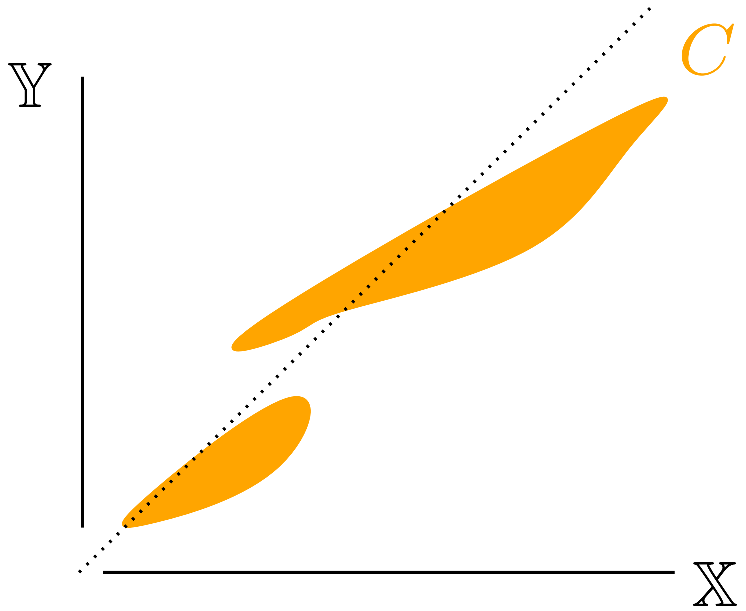
Need to be defined.

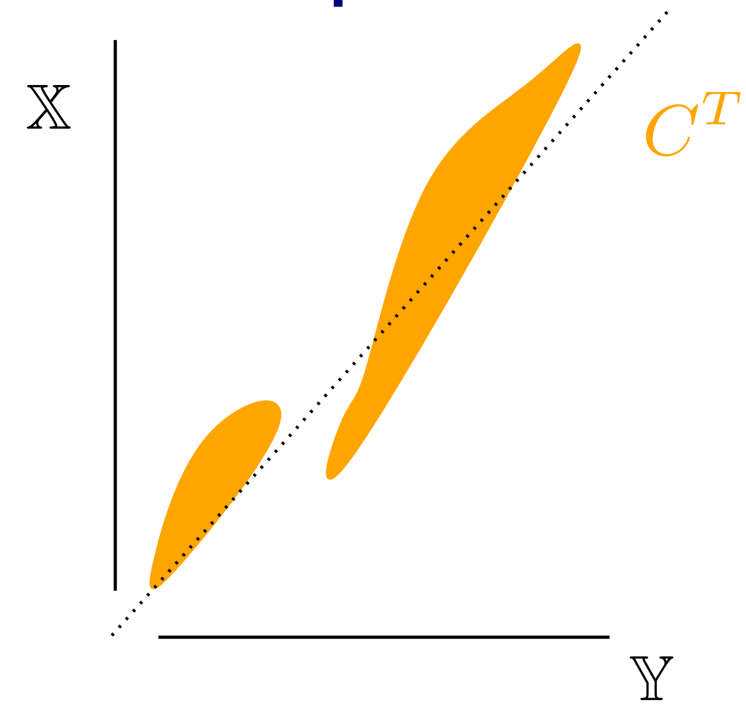
Multivalued maps and correspondences



A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ from a set $\mathbb{X}$ to a set $\mathbb{Y}$ is a subset of $\mathbb{X} \times \mathbb{Y}$, also denoted C, that projects surjectively onto $\mathbb{X}$ through the canonical projection $\pi_{\mathbb{X}} : \mathbb{X} \times \mathbb{Y} \to \mathbb{X}$. The image $C(\sigma)$ of a subset σ of $\mathbb{X}$ is the canonical projection onto $\mathbb{Y}$ of the preimage of σ through $\pi_{\mathbb{X}}$.

Multivalued maps and correspondences



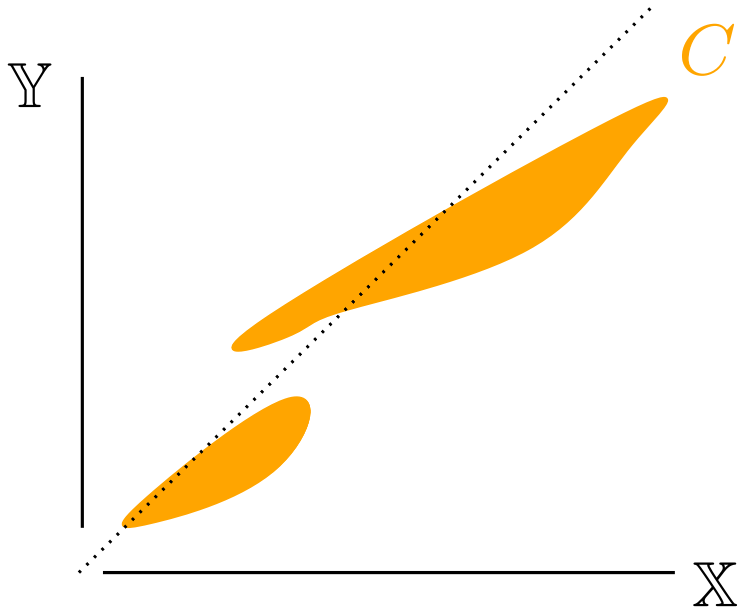


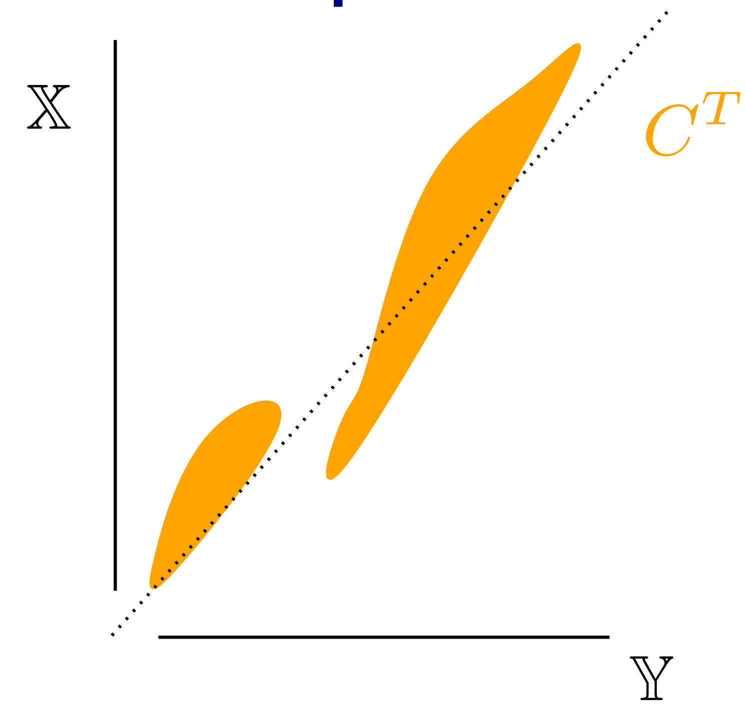
A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ from a set $\mathbb{X}$ to a set $\mathbb{Y}$ is a subset of $\mathbb{X} \times \mathbb{Y}$, also denoted C, that projects surjectively onto $\mathbb{X}$ through the canonical projection $\pi_{\mathbb{X}} : \mathbb{X} \times \mathbb{Y} \to \mathbb{X}$. The image $C(\sigma)$ of a subset σ of $\mathbb{X}$ is the canonical projection onto $\mathbb{Y}$ of the preimage of σ through $\pi_{\mathbb{X}}$.

The transpose of C, denoted C^T, is the image of C through the symmetry map $(x, y) \mapsto (y, x)$.

A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is a correspondence if C^T is also a multivalued map.

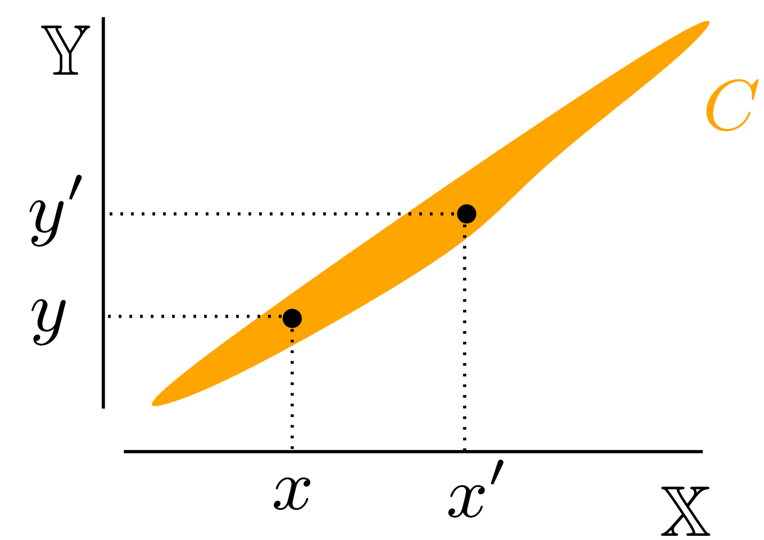
Multivalued maps and correspondences





A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ from a set $\mathbb{X}$ to a set $\mathbb{Y}$ is a subset of $\mathbb{X} \times \mathbb{Y}$, also denoted C, that projects surjectively onto $\mathbb{X}$ through the canonical projection $\pi_{\mathbb{X}} : \mathbb{X} \times \mathbb{Y} \to \mathbb{X}$. The image $C(\sigma)$ of a subset σ of $\mathbb{X}$ is the canonical projection onto $\mathbb{Y}$ of the preimage of σ through $\pi_{\mathbb{X}}$.

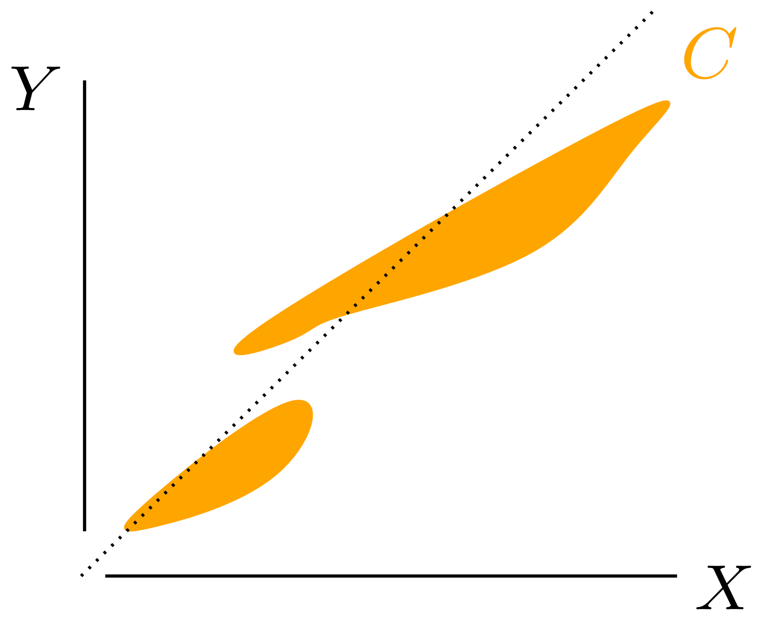
Example: ϵ-correspondence and Gromov-Hausdorff distance.

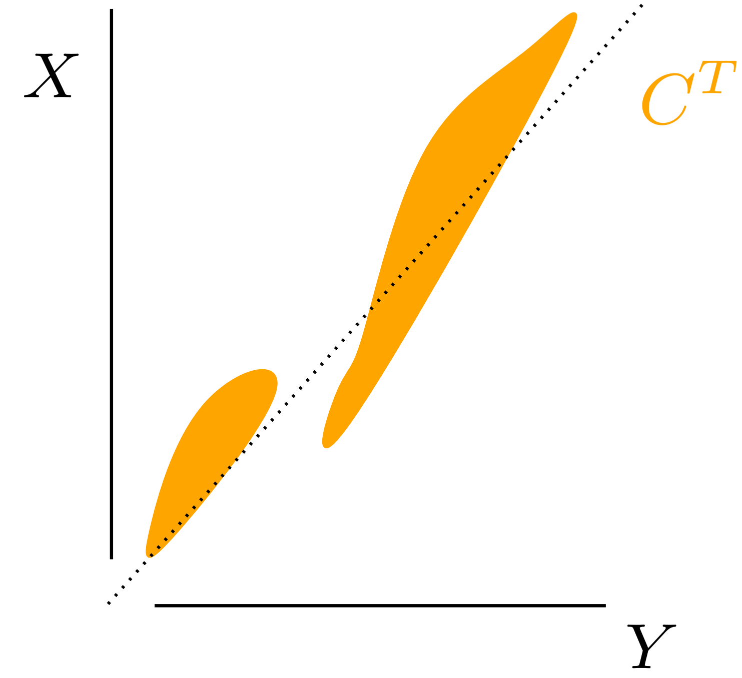


Let $(\mathbb{X}, \rho_{\mathbb{X}})$ and $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be compact metric spaces.
A correspondence $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is an ϵ-correspondence if $\forall (x, y), (x', y') \in C$, $|\rho_{\mathbb{X}}(x, x') - \rho_{\mathbb{Y}}(y, y')| \leq \varepsilon$.

$$d_{GH}(\mathbb{X}, \mathbb{Y}) = \frac{1}{2} \inf\{\varepsilon \geq 0 : \text{there exists an } \varepsilon\text{-correspondence between } \mathbb{X}\text{and } \mathbb{Y}\}$$

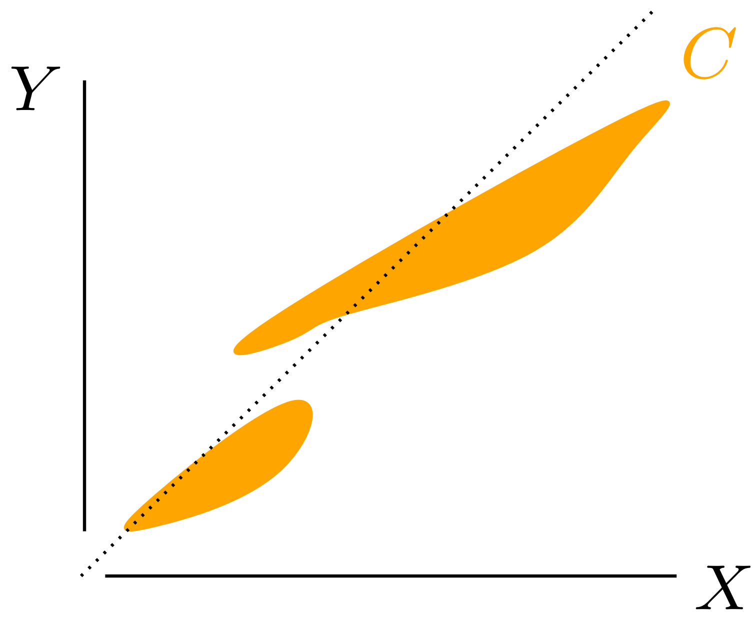
Multivalued simplicial maps

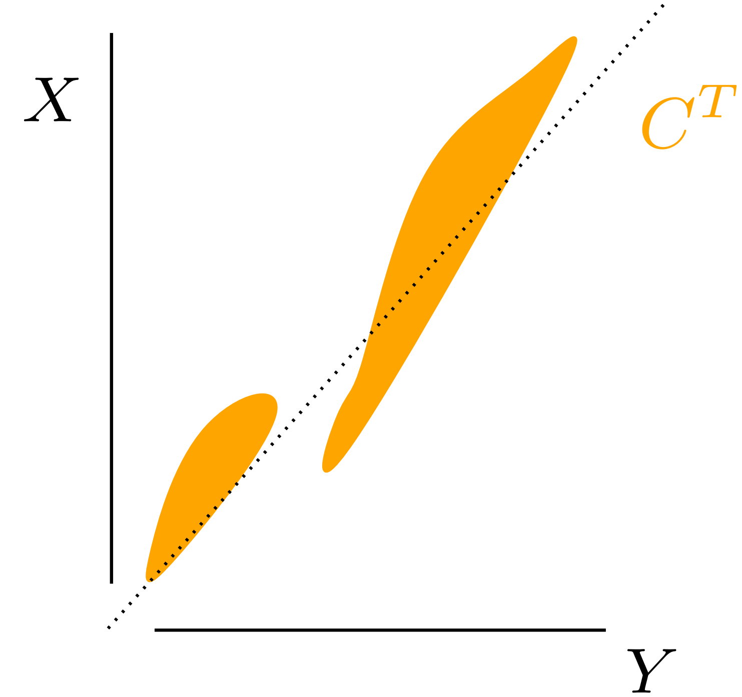




Let $\mathbb{S}$ and $\mathbb{T}$ be two filtered simplicial complexes with vertex sets $\mathbb{X}$ and $\mathbb{Y}$ respectively. A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is ε-simplicial from $\mathbb{S}$ to $\mathbb{T}$ if for any $a \in \mathbf{R}$ and any simplex $\sigma \in \mathbb{S}_a$, every finite subset of $C(\sigma)$ is a simplex of $\mathbb{T}_{a+\varepsilon}$.

Multivalued simplicial maps





Let $\mathbb{S}$ and $\mathbb{T}$ be two filtered simplicial complexes with vertex sets $\mathbb{X}$ and $\mathbb{Y}$ respectively. A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is ε-simplicial from $\mathbb{S}$ to $\mathbb{T}$ if for any $a \in \mathbf{R}$ and any simplex $\sigma \in \mathbb{S}_a$, every finite subset of $C(\sigma)$ is a simplex of $\mathbb{T}_{a+\varepsilon}$.

Proposition: Let $\mathbb{S}$, $\mathbb{T}$ be filtered complexes with vertex sets $\mathbb{X}$, $\mathbb{Y}$ respectively. If $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is a correspondence such that C and C^T are both ε-simplicial, then together they induce a canonical ε-interleaving between $\mathrm{H}(\mathbb{S})$ and $\mathrm{H}(\mathbb{T})$, the interleaving homomorphisms being $\mathrm{H}(C)$ and $\mathrm{H}(C^T)$.

The example of the Rips and Čech filtration

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}})$, $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\mathrm{Rips}(\mathbb{Y}))$ are ϵ-interleaved.

The example of the Rips and Čech filtration

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}})$, $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\mathrm{Rips}(\mathbb{Y}))$ are ϵ-interleaved.

Proof: Let $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ be a correspondence with distortion at most ϵ.
If $\sigma \in \mathrm{Rips}(\mathbb{X}, a)$ then $\rho_{\mathbb{X}}(x, x') \leq a$ for all $x, x' \in \sigma$.
Let $\tau \subseteq C(\sigma)$ be any finite subset.
For any $y, y' \in \tau$ there exist $x, x' \in \sigma$ s. t. $y \in C(x)$, $y' \in C(x')$ so

$$\rho_{\mathbb{Y}}(y, y') \leq \rho_{\mathbb{X}}(x, x') \leq a + \epsilon \text{ and } \tau \in \mathrm{Rips}(\mathbb{Y}, a + \epsilon)$$

$\Rightarrow C$ is ϵ-simplicial from $\mathrm{Rips}(\mathbb{X})$ to $\mathrm{Rips}(\mathbb{Y})$.
Symetrically, C^T is ϵ-simplicial from $\mathrm{Rips}(\mathbb{Y})$ to $\mathrm{Rips}(\mathbb{X})$.

The example of the Rips and Čech filtration

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}})$, $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\mathrm{Rips}(\mathbb{Y}))$ are ϵ-interleaved.

Proof: Let $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ be a correspondence with distortion at most ϵ.
If $\sigma \in \mathrm{Rips}(\mathbb{X}, a)$ then $\rho_{\mathbb{X}}(x, x') \leq a$ for all $x, x' \in \sigma$.
Let $\tau \subseteq C(\sigma)$ be any finite subset.
For any $y, y' \in \tau$ there exist $x, x' \in \sigma$ s. t. $y \in C(x)$, $y' \in C(x')$ so

$$\rho_{\mathbb{Y}}(y, y') \leq \rho_{\mathbb{X}}(x, x') \leq a + \epsilon \text{ and } \tau \in \mathrm{Rips}(\mathbb{Y}, a + \epsilon)$$

$\Rightarrow$ C is ϵ-simplicial from $\mathrm{Rips}(\mathbb{X})$ to $\mathrm{Rips}(\mathbb{Y})$.
Symetrically, C^T is ϵ-simplicial from $\mathrm{Rips}(\mathbb{Y})$ to $\mathrm{Rips}(\mathbb{X})$.

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}}), (\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{Y}))$ are ϵ-interleaved.

The example of the Rips and Čech filtration

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}})$, $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\mathrm{Rips}(\mathbb{Y}))$ are ϵ-interleaved.

Proof: Let $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ be a correspondence with distortion at most ϵ.
If $\sigma \in \mathrm{Rips}(\mathbb{X}, a)$ then $\rho_{\mathbb{X}}(x, x') \leq a$ for all $x, x' \in \sigma$.
Let $\tau \subseteq C(\sigma)$ be any finite subset.
For any $y, y' \in \tau$ there exist $x, x' \in \sigma$ s. t. $y \in C(x)$, $y' \in C(x')$ so

$$\rho_{\mathbb{Y}}(y, y') \leq \rho_{\mathbb{X}}(x, x') \leq a + \epsilon \text{ and } \tau \in \mathrm{Rips}(\mathbb{Y}, a + \epsilon)$$

$\Rightarrow C$ is ϵ-simplicial from $\mathrm{Rips}(\mathbb{X})$ to $\mathrm{Rips}(\mathbb{Y})$.
Symetrically, C^T is ϵ-simplicial from $\mathrm{Rips}(\mathbb{Y})$ to $\mathrm{Rips}(\mathbb{X})$.

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}}), (\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{Y}))$ are ϵ-interleaved.

Remark: Similar results for witness complexes (fixed landmarks)

Tameness of the Rips and Čech filtrations

Theorem: Let $\mathbb{X}$ be a precompact metric space. Then $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ are q-tame.

As a consequence dgm$(\mathrm{H}(\mathrm{Rips}(\mathbb{X})))$ and dgm$(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X})))$ are well-defined!

Tameness of the Rips and Čech filtrations

Theorem: Let $\mathbb{X}$ be a precompact metric space. Then $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ are q-tame.
As a consequence dgm$(\mathrm{H}(\mathrm{Rips}(\mathbb{X})))$ and dgm$(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X})))$ are well-defined!

Theorem: Let $\mathbb{X}, \mathbb{Y}$ be precompact metric spaces. Then

$$\mathrm{d_b}(\mathsf{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))), \mathsf{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{Y})))) \leq 2\mathrm{d_{GH}}(\mathbb{X}, \mathbb{Y}),$$

$$\mathrm{d_b}(\mathsf{dgm}(\mathrm{H}(\mathrm{Rips}(\mathbb{X}))), \mathsf{dgm}(\mathrm{H}(\mathrm{Rips}(\mathbb{Y})))) \leq 2\mathrm{d_{GH}}(\mathbb{X}, \mathbb{Y}).$$

Remark: The proofs never use the triangle inequality! The previous approch and results easily extend to other settings like, e.g. spaces endowed with a similarity measure.

Tameness of the Rips and Čech filtrations

Theorem: Let $\mathbb{X}$ be a precompact metric space. Then $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ are q-tame.
As a consequence dgm$(\mathrm{H}(\mathrm{Rips}(\mathbb{X})))$ and dgm$(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X})))$ are well-defined!

Theorem: Let $\mathbb{X}, \mathbb{Y}$ be precompact metric spaces. Then

$$\mathrm{d_b}(\mathsf{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))), \mathsf{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{Y})))) \leq 2\mathrm{d_{GH}}(\mathbb{X}, \mathbb{Y}),$$

$$\mathrm{d_b}(\mathsf{dgm}(\mathrm{H}(\mathrm{Rips}(\mathbb{X}))), \mathsf{dgm}(\mathrm{H}(\mathrm{Rips}(\mathbb{Y})))) \leq 2\mathrm{d_{GH}}(\mathbb{X}, \mathbb{Y}).$$

Remark: The proofs never use the triangle inequality! The previous approch and results easily extend to other settings like, e.g. spaces endowed with a similarity measure.

Theorem: Let $(\mathbb{M}, \rho)$ be homeomorphic to a locally finite simplicial complex, let $\mathbb{X}, \mathbb{Y} \subset \mathbb{M}$ be compact and let $\mathrm{Filt}(\mathbb{X})$ and $\mathrm{Filt}(\mathbb{Y})$ be the sublevel set filtrations of $\rho(\mathbb{X}, .)$ and $\rho(\mathbb{Y}, .)$. Then $H(\mathrm{Filt}(\mathbb{X}))$ and $H(\mathrm{Filt}(\mathbb{Y}))$ are q-tame [a] and

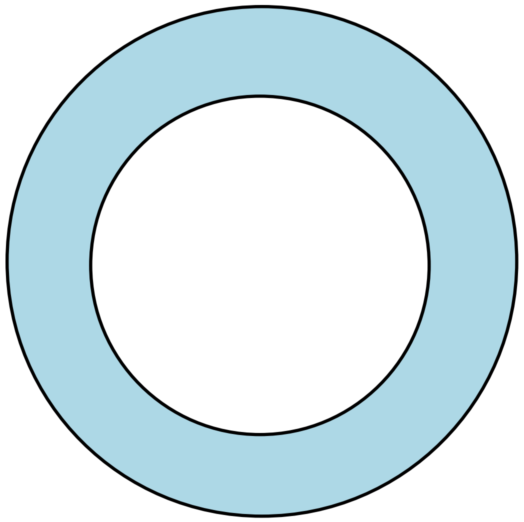
$$\mathrm{d_b}\,(\mathsf{dgm}(H(\mathrm{Filt}(\mathbb{X})), \mathsf{dgm}(H(\mathrm{Filt}(\mathbb{Y}))) \leq d_H(\mathbb{X}, \mathbb{Y})$$

---

[a] see also [Landi et al 2013]

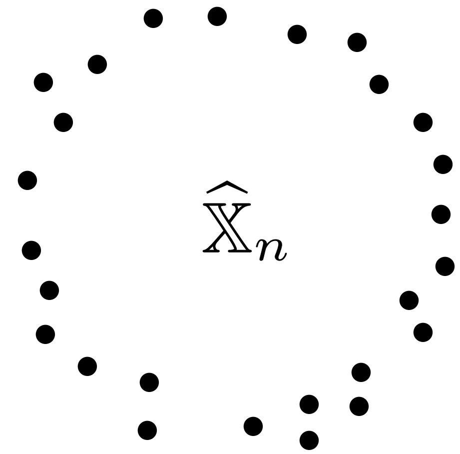
From stability to statistical properties

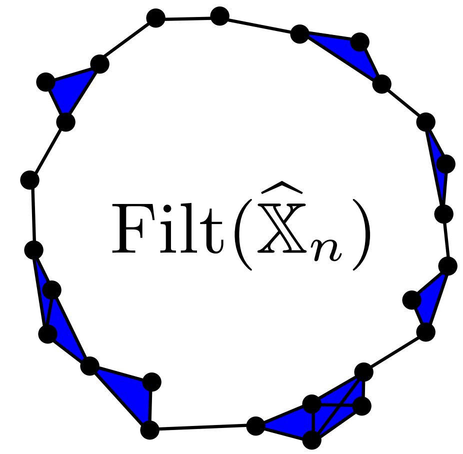
Statistical setting



$(\mathbb{M}, \rho)$ metric space
μ a probability measure with compact support $\mathbb{X}_\mu$.

$X_1, X_2, \cdots, X_n$
i.i.d. sampled according to μ.





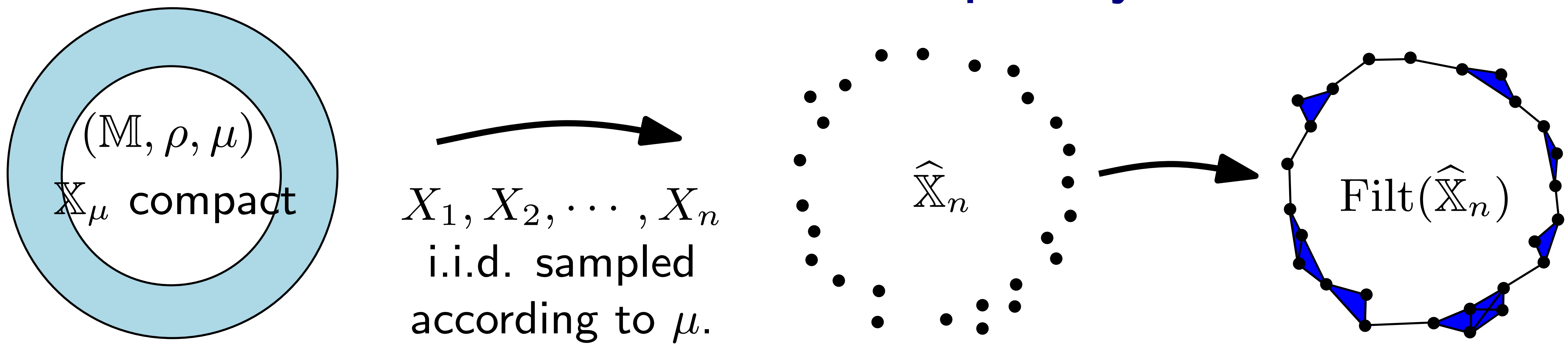
Examples:

- $\mathrm{Filt}(\widehat{\mathbb{X}}_n) = \mathrm{Rips}_\alpha(\widehat{\mathbb{X}}_n)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_n) = \check{\mathrm{C}}\mathrm{ech}_\alpha(\widehat{\mathbb{X}}_n)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_n) =$ sublevelset filtration of $\rho(., \mathbb{X}_\mu)$ (when $\mathbb{M}$ is a triangulable space).

From the previous stability theorem, for any $\varepsilon > 0$,

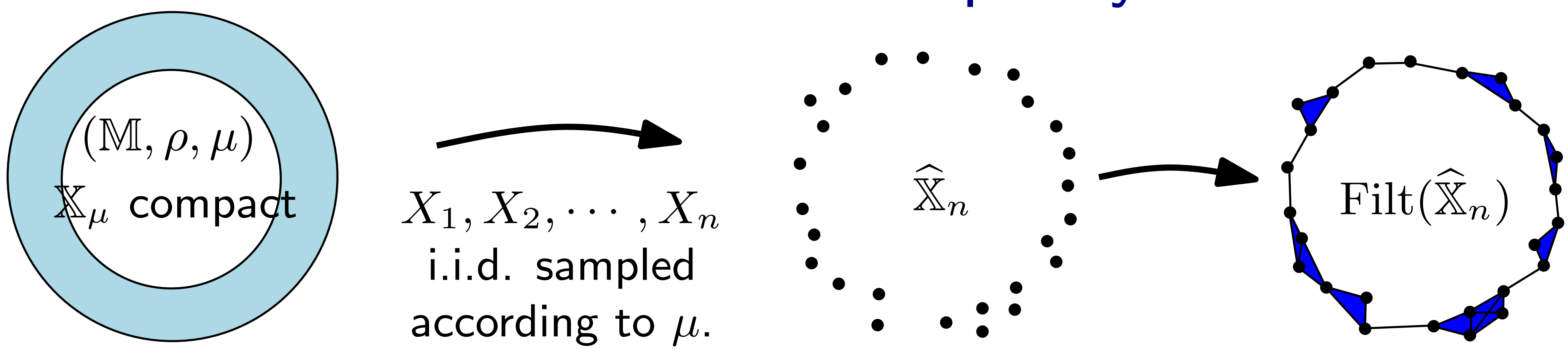
$$\mathbb{P}\left(\mathrm{d_b}\left(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathsf{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_n))\right) > \varepsilon\right) \leq \mathbb{P}\left(d_{GH}(\mathbb{X}_\mu, \widehat{\mathbb{X}}_n) > \frac{\varepsilon}{2}\right)$$

Concentration inequality



For $a, b > 0$, μ satisfies the (a, b)-standard assumption if for any $x \in \mathbb{X}_\mu$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

Concentration inequality



For $a, b > 0$, μ satisfies the (a, b)-standard assumption if for any $x \in \mathbb{X}_\mu$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

Theorem: If μ satisfies the (a, b)-standard assumption, then for any $\varepsilon > 0$:

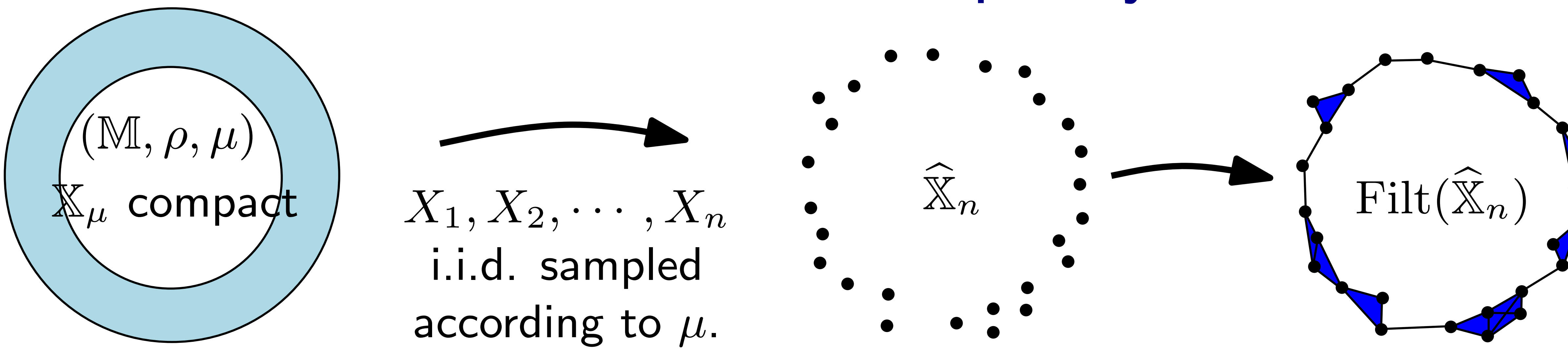
$$\mathbb{P}\left(\mathrm{d_b}\left(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathsf{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_n))\right) > \varepsilon\right) \leq \min\left(\frac{8^b}{a\varepsilon^b}\exp(-na\varepsilon^b), 1\right).$$

Moreover $\displaystyle \lim_{n\to\infty} \mathbb{P}\left(\mathrm{d_b}\left(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathsf{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_n))\right) \leq C_1 \left(\frac{\log n}{n}\right)^{1/b}\right) = 1.$

where C_1 is a constant only depending on a and b.

Remark: $\rightarrow$ Confidence intervals only depending on a and b (see also [Balakrishnan et al 2013] when $\mathbb{X}_\mu$ is a smooth manifold).

Concentration inequality



For $a, b > 0$, μ satisfies the (a, b)-standard assumption if for any $x \in \mathbb{X}_\mu$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

Sketch of proof:

1. Upperbound $\mathbb{P}\left(d_H(\mathbb{X}_\mu, \widehat{\mathbb{X}}_n) > \frac{\varepsilon}{2}\right)$.
2. (a, b) standard assumption $\Rightarrow$ an explicit upperbound for the covering number of $\mathbb{X}_\mu$ (by balls of radius $\varepsilon/2$).
3. Apply "union bound" argument.

Minimax rate of convergence

Let $\mathcal{P}(a, b, \mathbb{M})$ be the set of all the probability measures on the metric space $(\mathbb{M}, \rho)$ satisfying the (a, b)-standard assumption on $\mathbb{M}$:

Minimax rate of convergence

Let $\mathcal{P}(a, b, \mathbb{M})$ be the set of all the probability measures on the metric space $(\mathbb{M}, \rho)$ satisfying the (a, b)-standard assumption on $\mathbb{M}$:

Theorem:
Let $(\mathbb{M}, \rho)$ be a metric space and let $a > 0$ and $b > 0$. Then:

$$\sup_{\mu \in \mathcal{P}(a,b,\mathbb{M})} \mathbb{E}\left[\mathrm{d_b}(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathsf{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_n)))\right] \leq C \left(\frac{\ln n}{n}\right)^{1/b}$$

where the constant C only depends on a and b (not on $\mathbb{M}$!). Assume moreover that there exists a non isolated point x in $\mathbb{M}$ and consider any sequence $(x_n) \in (\mathbb{M} \setminus \{x\})^{\mathbb{N}}$ such that $\rho(x, x_n) \leq (an)^{-1/b}$. Then for any estimator $\widehat{\mathsf{dgm}}_n$ of $\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu))$:

$$\liminf_{n \to \infty} \rho(x, x_n)^{-1} \sup_{\mu \in \mathcal{P}(a,b,\mathbb{M})} \mathbb{E}\left[\mathrm{d_b}(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \widehat{\mathsf{dgm}}_n)\right] \geq C'$$

where C' is an absolute constant.

Remark: we can obtain slightly better bounds if $\mathbb{X}_\mu$ is a submanifold of $\mathbb{R}^D$ - see [Genovese, Perone-Pacifico,Verdinelli, Wasserman 2011, 2012]

Lecam's Lemma

Lemma:
Let $\mathcal{P}$ be a set of proba distributions. For $\mu \in \mathcal{P}$, let $\theta(\mu)$ take values in a metric space $(\mathbb{X}, \rho_{\mathbb{X}})$. Let μ_0 and μ_1 in $\mathcal{P}$ be any pair of distributions. Let $X_1, \ldots, X_n$ be drawn i.i.d. from some $\mu \in \mathcal{P}$. Let $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ be any estimator of $\theta(\mu)$, then

$$\sup_{\mu \in \mathcal{P}} \mathbb{E}_{\mu^n} \rho_{\mathbb{X}}(\theta, \hat{\theta}) \geq \frac{1}{8} \rho_{\mathbb{X}} \left(\theta(\mu_0), \theta(\mu_1)\right) \left[1 - \mathrm{TV}(\mu_0, \mu_1)\right]^{2n}.$$

where $\mathrm{TV}(\mu_0, \mu_1) = \sup_{B \in \mathcal{B}} |\mu_0(B) - \mu_1(B)|$.

Lecam's Lemma

Lemma:
Let $\mathcal{P}$ be a set of proba distributions. For $\mu \in \mathcal{P}$, let $\theta(\mu)$ take values in a metric space $(\mathbb{X}, \rho_{\mathbb{X}})$. Let μ_0 and μ_1 in $\mathcal{P}$ be any pair of distributions. Let $X_1, \ldots, X_n$ be drawn i.i.d. from some $\mu \in \mathcal{P}$. Let $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ be any estimator of $\theta(\mu)$, then

$$\sup_{\mu \in \mathcal{P}} \mathbb{E}_{\mu^n} \rho_{\mathbb{X}}(\theta, \hat{\theta}) \geq \frac{1}{8} \rho_{\mathbb{X}}\left(\theta(\mu_0), \theta(\mu_1)\right) \left[1 - \mathrm{TV}(\mu_0, \mu_1)\right]^{2n}.$$

where $\mathrm{TV}(\mu_0, \mu_1) = \sup_{B \in \mathcal{B}} |\mu_0(B) - \mu_1(B)|$.

In our case:

$\mathcal{P} = \mathcal{P}(a, b, \mathbb{M})$, $(\mathbb{X}, \rho_{\mathbb{X}})$ is the space of persistence diagrams with $\rho_{\mathbb{X}} = d_B$ and $\theta(\mu) = \mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu))$.

$\mu_0 = \delta_x$ the Dirac mass at x and $\mu_1 = \frac{1}{n}\delta_{x_n} + (1 - \frac{1}{n})\mu_0$ (they belong to $\mathcal{P}$).

$\mathrm{TV}(\mu_0, \mu_1) = \frac{2}{n}$, so $[1 - \mathrm{TV}(\mu_0, \mu_1)]^{2n} \to e^{-4}$ as $n \to \infty$.

$\mathrm{d_b}(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_0)), \mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_1))) = \rho_{\mathbb{M}}(x, x_n)/2$

References:

- F. Chazal, M. Glisse, C. Labruère, B. Michel, Optimal rates of convergence for persistence diagrams in Topological Data Analysis, http://arxiv.org/abs/1305.6239, May 2013.
- F. Chazal, V. de Silva, S. Oudot, Persistence Stability for Geometric complexes, arXiv:1207.3885, July 2012.
- F. Chazal, V. de Silva, M. Glisse, S. Oudot, The Structure and Stability of Persistence Modules, arXiv:1207.3674, July 2012.