A limit theorem for persistence diagrams of random filtered complexes built over marked point processes

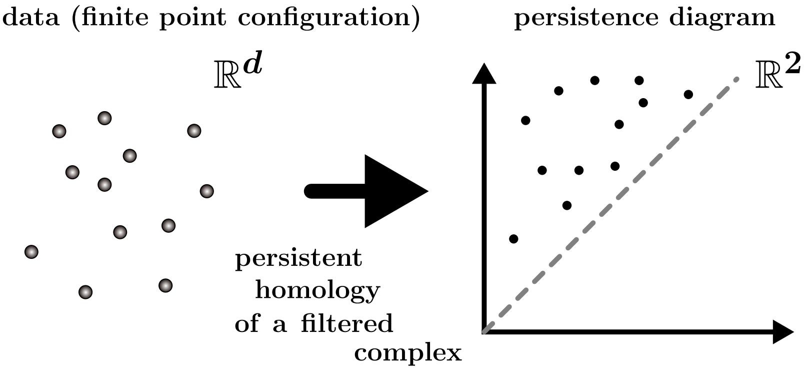
Kiyotaka Suzaki

Kumamoto University

Sep. 3 2019

joint work with Tomoyuki Shirai (Kyushu University)

What are persistence diagrams?



What are persistence diagrams?

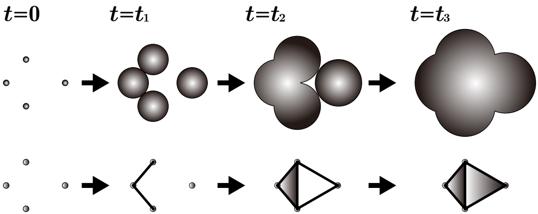


Figure: Čech complex built over data

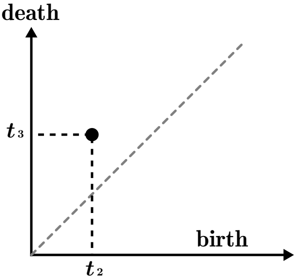


Figure: 1st persistence diagram

Application to characterizing atomic structures

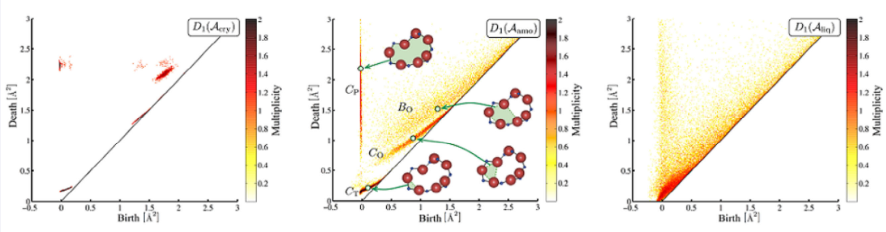


Figure:

1st persistence diagrams for SiO2
(left: crystal, middle: glass, right: liquid)

(Hiraoka Laboratory, https://sites.google.com/view/hiraoka-lab-en/research/applied-research/tda-on-amorphous-structures?authuser=0)

Law of large numbers for persistence diagrams

Theorem (Hiraoka, Shirai, and Trinh (2018))

$\boldsymbol{\Phi}$: stationary point process on $\mathbb{R}^d$ having all finite moment.

$\boldsymbol{\xi_{q,L}}$: persistence diagram of "$\boldsymbol{\kappa}$-filtered complex"
built over random data $\boldsymbol{\Phi} \cap [-L/2, L/2)^d$
($\boldsymbol{\xi_{q,L}}$ can be regarded as a point process on $\mathbb{R}^2$)

Then

$$\frac{1}{L^d}\mathbb{E}[\xi_{q,L}] \xrightarrow{v} \nu_q \quad \text{as} \quad L \to \infty.$$

Furthermore if $\boldsymbol{\Phi}$ is ergodic, then

$$\frac{1}{L^d}\xi_{q,L} \xrightarrow{v} \nu_q \quad \text{as} \quad L \to \infty \text{ a.s.}$$

Today's talk

input data $\longrightarrow$ filtered complex $\longrightarrow$ persistence diagram

In this talk

Input data	$\rightsquigarrow$	**marked point process**		
$\boldsymbol{\kappa}$**-filtered complex**	$\rightsquigarrow$	**that of MPP**	$\Rightarrow$	**LLN for PDs**
$[-L/2, L/2)^d$	$\rightsquigarrow$	**convex set**		

κ-filtered complex built over marked point set

S : set, $\quad \mathcal{F}(S) = \{A \subset S : 0 < \#A < \infty\}$,

$\mathbb{M}$: locally compact, 2nd countable Hausdorff space,

$\kappa : \mathcal{F}(\mathbb{R}^d \times \mathbb{M}) \to [0, \infty)$ satisfying:

(K1) $\kappa(A) \le \kappa(B)$ if $A \subset B$.

(K2) κ is invariant under the translations on $\mathbb{R}^d$, i.e., $\kappa(T_a A) = \kappa(A)$ for any $a \in \mathbb{R}^d$ and $A \in \mathcal{F}(\mathbb{R}^d \times \mathbb{M})$, where $T_a : (x, m) \mapsto (x + a, m)$.

(K3) There exists an increasing function $\rho : [0, \infty) \to [0, \infty)$ such that $|x - y| \le \rho(\kappa(\{(x, m), (y, n)\}))$ for all $(x, m), (y, n) \in \mathbb{R}^d \times \mathbb{M}$.

κ-filtered complex built over marked point set

$\pi : \mathbb{R}^d \times \mathbb{M} \ni (x, m) \mapsto x \in \mathbb{R}^d$,
$\widetilde{\Xi} \in \mathcal{F}(\mathbb{R}^d \times \mathbb{M})$ is a simple marked point set
$\overset{\text{def}}{\Leftrightarrow}$ for any $x \in \mathbb{R}^d$, $\#(\widetilde{\Xi} \cap \pi^{-1}\{x\}) = 0$ or 1.
$\Xi = \pi(\widetilde{\Xi})$,

$$\widetilde{\Xi} = \{(x_0, m_0), (x_1, m_1), \ldots, (x_q, m_q)\}$$
$$\simeq \Xi = \{x_0, x_1, \ldots, x_q\} \subset \mathbb{R}^d \ \& \ \{m_0, m_1, \ldots, m_q\} \subset \mathbb{M}$$

finite point data & marks

π induces the bijection

$$\mathcal{F}(\widetilde{\Xi}) \ni \widetilde{\sigma} \mapsto \sigma \in \mathcal{F}(\Xi).$$

κ-filtered complex built over marked point set

Given a simple marked point set $\widetilde{\Xi}$, we define a filtration

$$\mathbb{K}(\widetilde{\Xi}) = \{K(\widetilde{\Xi}, t)\}_{t \geq 0}$$

of simplicial complexes from $\tilde{\Xi}$ by

$$K(\widetilde{\Xi}, t) = \{\sigma \subset \Xi \,:\, \kappa(\widetilde{\sigma}) \leq t\}.$$

Remarks

- $\kappa(\tilde{\sigma})$ = the birth time of a simplex σ in the filtration $\mathbb{K}(\tilde{\Xi})$.
- **(K3)** $\rightsquigarrow$ $\quad \sigma \in K(\widetilde{\Xi}, t) \Rightarrow \operatorname{diam} \sigma \leq \rho(t)$

We call $\mathbb{K}(\widetilde{\Xi})$ a κ-filtered complex built over $\widetilde{\Xi}$.

Examples of κ-filtered complex

(i) $\mathbb{M} = [0, R]$ (the set of radii $\leq R$),

$$\kappa(\widetilde{\sigma}) = \inf_{w \in \mathbb{R}^d} \max_{(x,r) \in \widetilde{\sigma}} \{|x - w| - r\}^+$$

$$\kappa(\widetilde{\sigma}) \leq t \quad \Leftrightarrow \quad \bigcap_{(x,r) \in \widetilde{\sigma}} \overline{B}(x\,;\,t + r) \neq \emptyset$$

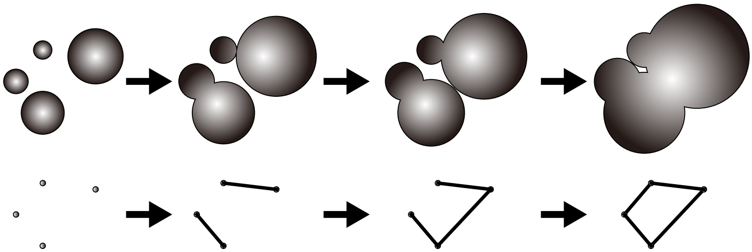


Figure: Čech complex of balls with various sizes

Examples of κ-filtered complex (cont.)

(ii) $\mathbb{M} = \{t, t^n, e^t - 1, \log(1+t), \ldots,$ etc. $\}$
(a finite family of continuous strictly increasing functions),

$$\kappa(\widetilde{\sigma}) = \inf_{w \in \mathbb{R}^d} \max_{(x, r(\cdot)) \in \widetilde{\sigma}} r^{-1}(|x - w|)$$

$$\kappa(\widetilde{\sigma}) \le t \iff \bigcap_{(x, r(\cdot)) \in \widetilde{\sigma}} \overline{B}(x\,;\, r(t)) \neq \emptyset$$

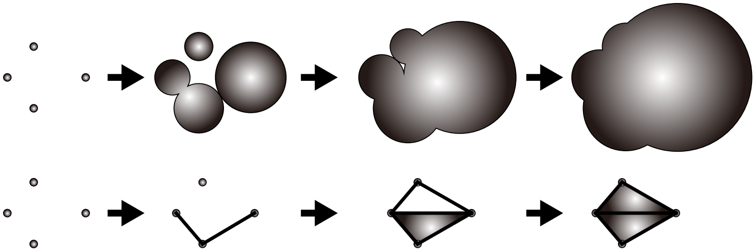


Figure: Čech complex of balls with various growths

Examples of $\boldsymbol{\kappa}$-filtered complex (cont.)

(iii)′ $\mathbb{M} = \{$ball, solid convex polytope, solid ellipsoid, $\ldots,$ etc. $\}$
Similarly, we can define a $\boldsymbol{\kappa}$ such that the corresponding $\boldsymbol{\kappa}$-filtered complex is

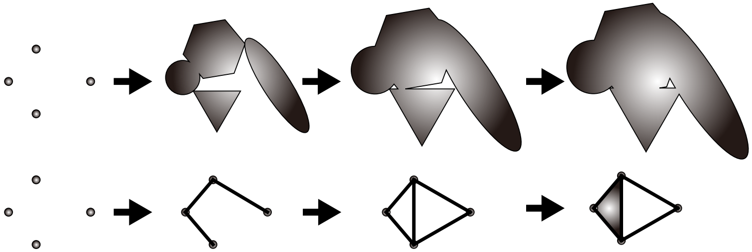


Figure: Čech complex of sets with various shapes

Persistent homology

$\mathbb{K}(\widetilde{\Xi}) = \{K(\widetilde{\Xi}, t)\}_{t \geq 0}$: κ-filtered complex built over $\widetilde{\Xi}$,

$$H_q(K(\widetilde{\Xi}, t)) = q \text{ th homology module of } K(\widetilde{\Xi}, t) \text{ on a field } \mathbb{F}$$

$$= \begin{cases} \text{connected components} & q = 0 \\ \text{rings} & q = 1 \\ \text{cavities} & q = 2 \\ \quad \vdots & \quad \vdots \end{cases}$$

- For $r \leq s$, $K(\widetilde{\Xi}, r) \hookrightarrow K(\widetilde{\Xi}, s)$ induces the linear map

$$\iota_r^s : H_q(K(\widetilde{\Xi}, r)) \to H_q(K(\widetilde{\Xi}, s)).$$

Persistence diagram

- $H_q(\mathbb{K}(\widetilde{\Xi})) = (\{H_q(K(\widetilde{\Xi}, t))\}_{t \geq 0}, \{\iota_r^s\}_{r \leq s})$
  : **q-th persistent homology (module)**

$$H_q(\mathbb{K}(\widetilde{\Xi})) \simeq \bigoplus_{i=1}^{n_q} I(b_i, d_i),$$

where $I(b_i, d_i) = (\{U_t\}_{t \geq 0}, \{f_r^s\}_{r \leq s})$

$$U_t = \begin{cases} \mathbb{F} & b_i \leq t < d_i, \\ \{0\} & \text{otherwise}, \end{cases}$$

and $f_r^s = \mathrm{id}_{\mathbb{F}}$ for $b_i \leq r \leq s < d_i$.

- $I(b_i, d_i)$ means a "q-dimensional hole" appears at $t = b_i$, persists $b_i \leq t < d_i$ and disappears at $t = d_i$.

Persistence diagrams as counting measures

- $D_q(\mathbb{K}(\widetilde{\Xi})) = \{(b_i, d_i) \in \Delta : i = 1, 2, \ldots, n_q\}$
  : **q-th persistence diagram** of $\mathbb{K}(\widetilde{\Xi})$,
  where
  $\Delta = \{(x, y) \in [0, \infty] \times [0, \infty] : 0 \leq x < y \leq \infty\}$.

Remark

We deal with $D_q(\mathbb{K}(\widetilde{\Xi}))$ as the counting measure

$$\xi_q(\mathbb{K}(\widetilde{\Xi})) = \sum_{(b_i, d_i) \in D_q(\mathbb{K}(\widetilde{\Xi}))} \delta_{(b_i, d_i)}.$$

Marked point process

$\widetilde{\Phi}$ is a **marked point process** on $\mathbb{R}^d$ with marks in $\mathbb{M}$
$\overset{\text{def}}{\Leftrightarrow}$ $\widetilde{\Phi}$ is a point process on $\mathbb{R}^d \times \mathbb{M}$ such that for any $x \in \mathbb{R}^d$,
$\widetilde{\Phi}(\pi^{-1}\{x\}) = 0$ or 1 a.s.

- $\widetilde{\Phi} = \sum \delta_{(X_i, M_i)}$,
  $\{(X_i, M_i)\}$ are $\mathbb{R}^d \times \mathbb{M}$-valued random variables.
- For a bounded Borel $A \subset \mathbb{R}^d$,
  $\{(X_i, M_i)\} \cap (A \times \mathbb{M})$ is a (random) simple marked point set.
- $\Phi = \sum \delta_{X_i}$: the ground process of $\widetilde{\Phi}$
- Φ has all finite moment
  $\overset{\text{def}}{\Leftrightarrow}$ for all compact $A \subset \mathbb{R}^d$ and $p \geq 1$, $\mathbb{E}[\Phi(A)^p] < \infty$

Translations on $\mathbf{Conf}(\mathbb{R}^d \times \mathbb{M})$

For $a \in \mathbb{R}^d$,

$$T_a : \mathbb{R}^d \times \mathbb{M} \ni (x, m) \mapsto (x + a, m) \in \mathbb{R}^d \times \mathbb{M}$$

$$\Downarrow$$

$$(T_a)_* : \mathbf{Conf}(\mathbb{R}^d \times \mathbb{M}) \ni \mu \mapsto \mu \circ T_a^{-1} \in \mathbf{Conf}(\mathbb{R}^d \times \mathbb{M}).$$

Stationarity and Ergodicity of $\widetilde{\boldsymbol{\Phi}}$

$$\widetilde{\boldsymbol{\Phi}} \text{ is stationary } \overset{\text{def}}{\Leftrightarrow} \boldsymbol{P} \circ ((\boldsymbol{T_a})_* \widetilde{\boldsymbol{\Phi}})^{-1} = \boldsymbol{P} \circ \widetilde{\boldsymbol{\Phi}}^{-1} \text{ for any } \boldsymbol{a} \in \mathbb{R}^{\boldsymbol{d}}$$

$$\widetilde{\boldsymbol{\Phi}} \text{ is ergodic } \overset{\text{def}}{\Leftrightarrow} \begin{cases} \quad \boldsymbol{I} \in \boldsymbol{\mathcal{B}}(\mathbf{Conf}(\mathbb{R}^{\boldsymbol{d}} \times \mathbb{M})), \\ \quad \boldsymbol{P} \circ \widetilde{\boldsymbol{\Phi}}^{-1}(\boldsymbol{I} \boldsymbol{\Delta} (\boldsymbol{T_a})_* \boldsymbol{I}) = \mathbf{0} \text{ for any } \boldsymbol{a} \in \mathbb{R}^{\boldsymbol{d}} \\ \Rightarrow \boldsymbol{P} \circ \widetilde{\boldsymbol{\Phi}}^{-1}(\boldsymbol{I}) \in \{\mathbf{0}, \mathbf{1}\} \end{cases}$$

Convex averaging net

$(\mathcal{N}, \leq)$: linearly ordered set

$\mathcal{A} = \{A_n\}_{n \in \mathcal{N}}$: a family of bounded Borel sets in $\mathbb{R}^d$

is a **convex averaging net** if

(i) A_n is convex for each $n \in \mathcal{N}$,

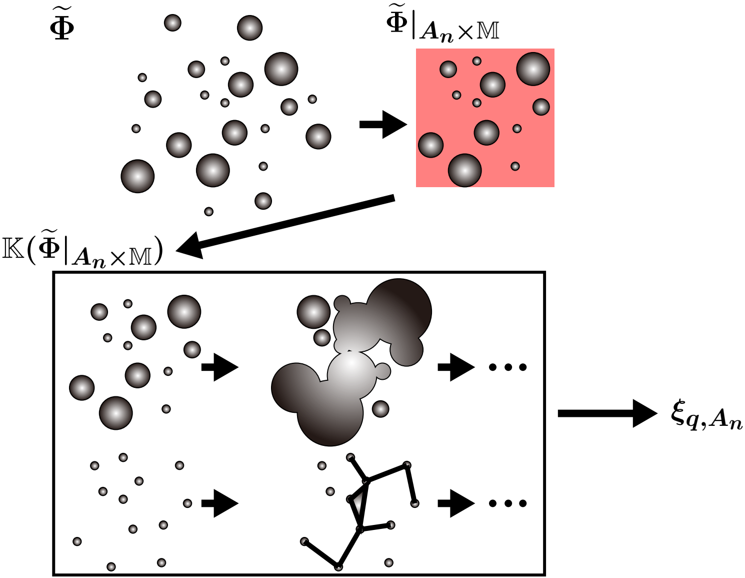
(ii) $A_n \subset A_m$ for $n \leq m$, and

(iii) $\sup_{n \in \mathcal{N}} r(A_n) = \infty$,
where $r(A) = \sup\{r > 0 \ : \ A \text{ contains a ball of radius } r\}$.

We put

$$\xi_{q,A_n} = \xi_q(\mathbb{K}(\widetilde{\Phi}|_{A_n \times \mathbb{M}})),$$

i.e.,


$\widetilde{\Phi}$
$\widetilde{\Phi}|_{A_n \times \mathbb{M}}$
$\mathbb{K}(\widetilde{\Phi}|_{A_n \times \mathbb{M}})$
ξ_{q,A_n}

Main result

Theorem

Let $\widetilde{\Phi}$ be a stationary marked point process and suppose its ground process Φ has all finite moments. Then for any nonnegative integer q, there exists a Radon measure ν_q on Δ such that for any convex averaging net $\mathcal{A} = \{A_n\}_{n \in \mathcal{N}}$ in $\mathbb{R}^d$,

$$\frac{1}{\ell(A_n)} \mathbb{E}[\xi_{q,A_n}] \xrightarrow{v} \nu_q \quad \text{as} \quad n \to \infty,$$

where ℓ is the d-dimensional Lebesgue measure. Furthermore if $\widetilde{\Phi}$ is ergodic, then

$$\frac{1}{\ell(A_n)} \xi_{q,A_n} \xrightarrow{v} \nu_q \quad \text{as} \quad n \to \infty \text{ a.s.}$$

Sketch of the Proof

- We extend ideas in Hiraoka, Shirai, and Trinh (2018) to marked point processes.
- $$\begin{aligned}\beta_q^{r,s}(\mathbb{K}(\widetilde{\Xi})) &= \begin{array}{c}\text{the rank of}\\ \iota_r^s : H_q(K(\widetilde{\Xi},r)) \to H_q(K(\widetilde{\Xi},s))\end{array}\\ &= \xi_q(\mathbb{K}(\widetilde{\Xi}))([0,r)\times(s,\infty])\end{aligned}$$

  is called the q**th** (r,s)**-persistent Betti number** of $\mathbb{K}(\widetilde{\Xi})$.
  The LLN for $\beta_q^{r,s}$ + general theory of Radon measures
  $\Rightarrow$ Main theorem is proved.

Lemma

Let $\mathbb{K}^1 = \{K^1_t\}_{t\geq 0}$ *and* $\mathbb{K}^2 = \{K^2_t\}_{t\geq 0}$ *be filtrations with* $K^1_t \subset K^2_t$ *for* $t \geq 0$*. Then,*

$$\begin{aligned} &|\beta^{r,s}_q(\mathbb{K}^1) - \beta^{r,s}_q(\mathbb{K}^2)| \\ &\leq \sum_{j=q,q+1} \# K^2_{s,j} \setminus K^1_{s,j} + \#\{\sigma \in K^1_{s,j} \setminus K^1_{r,j} \,:\, t^{(2)}_\sigma \leq r\}, \end{aligned}$$

where $K^i_{s,j}$ *is the set of* j*-simplices in* K^i_s*, and* $t^{(i)}_\sigma$ *is the birth time of* σ *in* $\mathbb{K}^i$*,* $i = 1, 2$*.*

$$\beta^{r,s}_{q,A_n} = \beta^{r,s}_q(\mathbb{K}(\widetilde{\Phi}|_{A_n \times \mathbb{M}}))$$

Theorem

Let $\widetilde{\Phi}$ be a stationary marked point process and suppose its ground process Φ has all finite moments. Then, for any $0 \leq r \leq s < \infty$ and nonnegative integer q, there exists a nonnegative number $\hat{\beta}^{r,s}_q$ such that for any convex averaging net $\mathcal{A} = \{A_n\}_{n \in \mathcal{N}}$ in $\mathbb{R}^d$,

$$\frac{1}{\ell(A_n)}\mathbb{E}[\beta^{r,s}_{q,A_n}] \to \hat{\beta}^{r,s}_q \qquad \text{as} \quad n \to \infty.$$

Furthermore, if $\widetilde{\Phi}$ is ergodic, then

$$\frac{1}{\ell(A_n)}\beta^{r,s}_{q,A_n} \to \hat{\beta}^{r,s}_q \quad \text{as} \quad n \to \infty \text{ a.s.}$$