Statistical Analysis of Persistent Homology

Genki Kusano (Tohoku University, D1)

Topology and Computer 2016, Oct 28 @ Akita.

Collaborators：

Kenji Fukumizu (The Institute of Statistical Mathematics）

Yasuaki Hiraoka (Tohoku University, AIMR)

Persistence weighted Gaussian kernel for topological data analysis.
Proceedings of the 33rd ICML, pp. 2004–2013, 2016

Self introduction

- Interests : Applied topology, topological data analysis

  B3 : Homology group, homological algebra

  B4 : Persistent homology, computational homology

  M1 : Applied topology to sensor network

  “**Relative interleavings and applications to sensor networks**”, JJIAM, 33(1),99-120, 2016.

  M2 : Statistics, machine learning, kernel methods

  “**Persistence weighted Gaussian kernel for topological data analysis**”, ICML, pp. 2004–2013, 2016.

  D1(now) : Time series analysis, dynamics, information geometry, …

- Announcement : Joint Mathematics Meetings, January 4, 2017, Atlanta

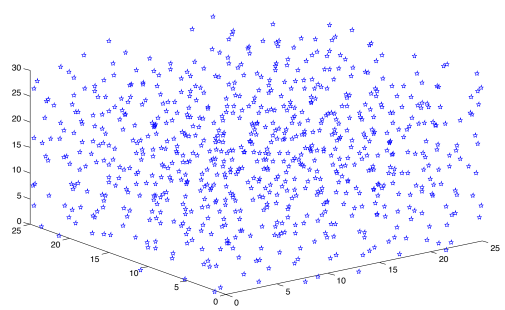
  ★Statistical Methods in Computational Topology and Applications

  Sheaves in Topological Data Analysis

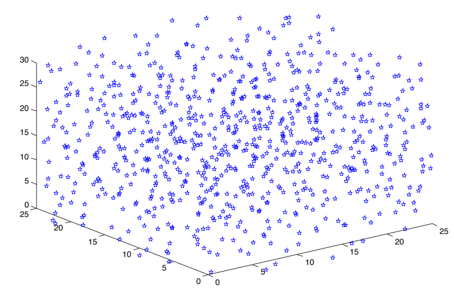
Motivation of this work

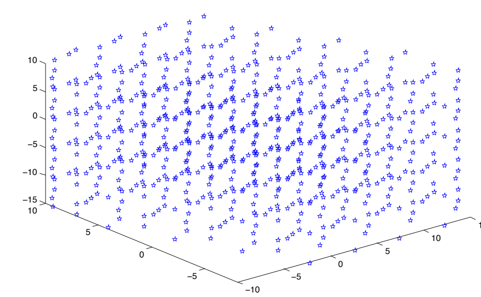
Topological Data Analysis (TDA, 位相的データ解析)

Mathematical methods for characterizing “shapes of data”



? ⟷





Liquid Glass Solid

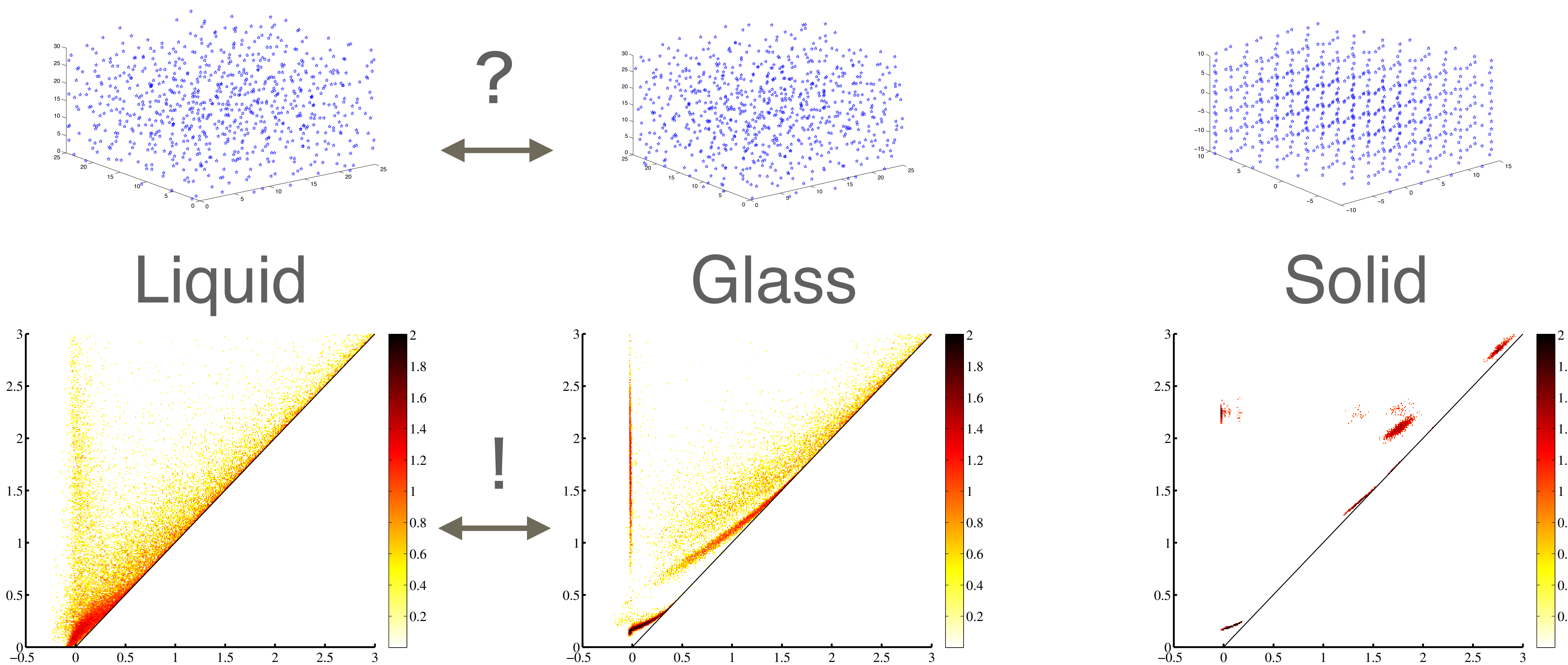
Atomic configurations of liquid, glass, and solid state of silica ($\mathrm{SiO_2}$, silica — composed of silicon and oxygen)

At the configuration level, it is difficult to distinguish liquid and glass state.

Motivation of this work

Persistent homology / Persistence diagram

Topological descriptor of data

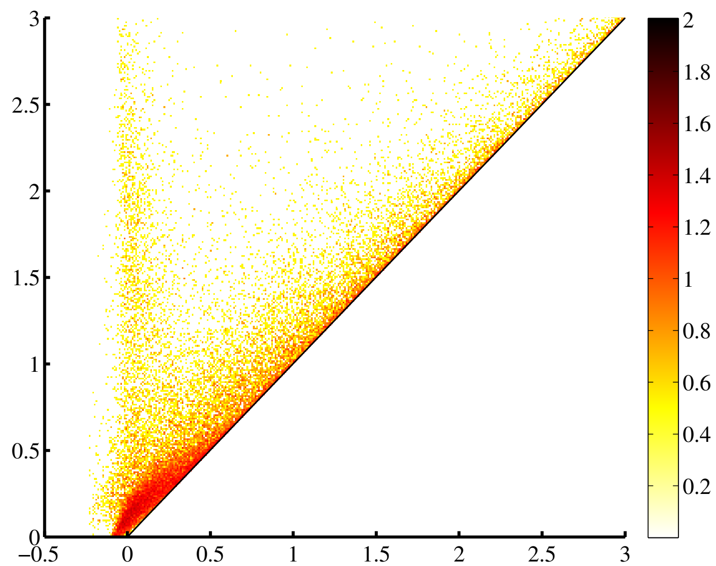


Y. Hiraoka et al., Hierarchical structures of amorphous solids characterized by persistent homology, PNAS, 113(26):7035–7040, 2016.

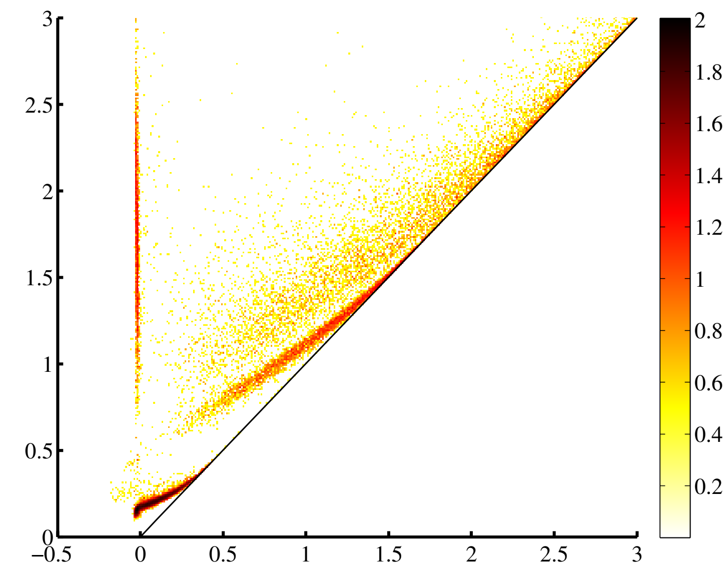
Motivation of this work

Classification problem

Liquid



Glass

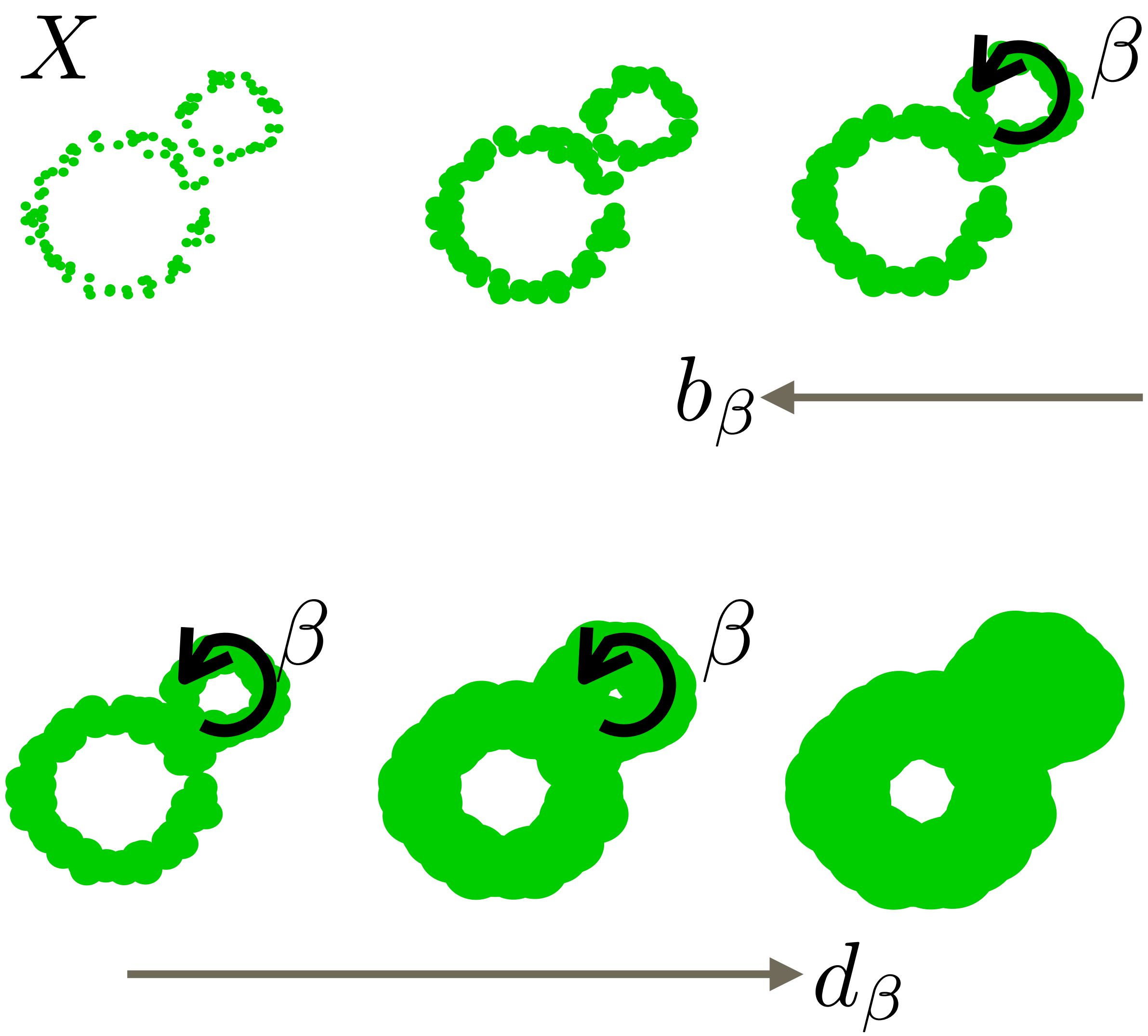


Q. Can we distinguish them mathematically?

A. Make a **statistical framework** for persistence diagrams

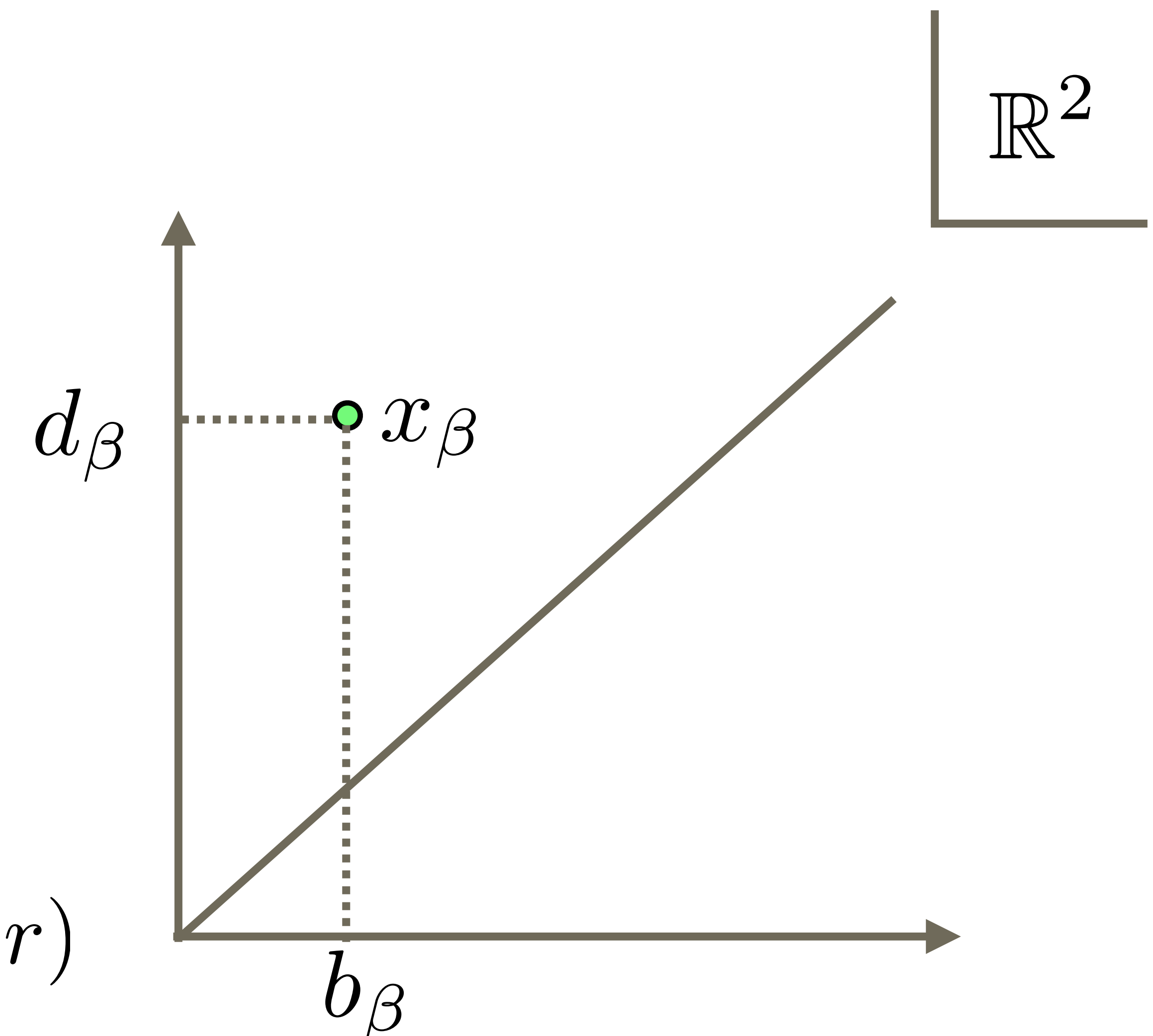
Section 1
What is a persistence diagram

Persistence diagram

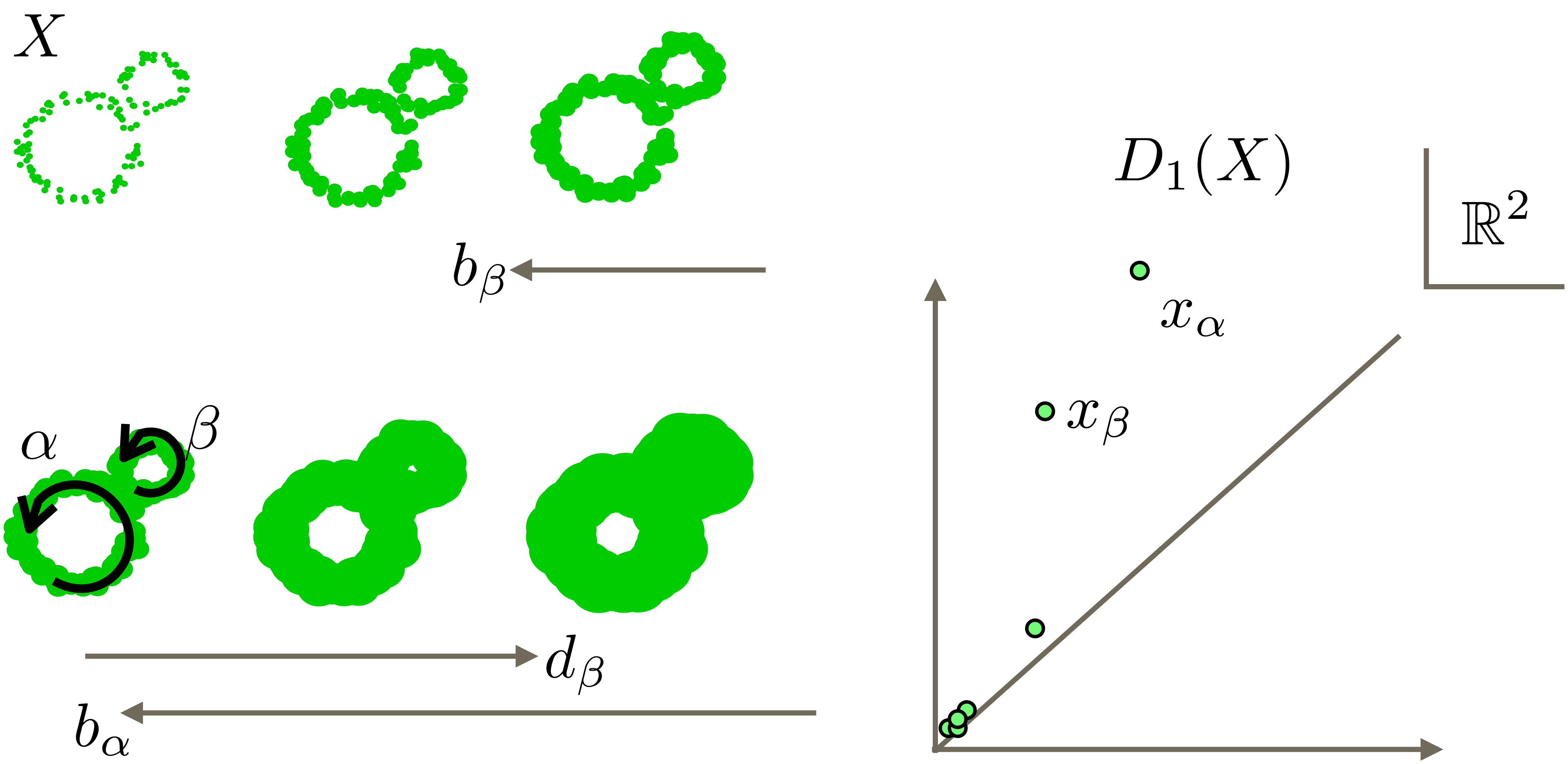


$$X_r = \bigcup_{x_i \in X} B(x_i; r)$$

$$B(x; r) = \{z \mid d(x,z) \le r\}$$



Persistence diagram



$$D_1(X) = \{x_\alpha, x_\beta, x_\gamma, \cdots \mid x_a = (b_\alpha, d_\alpha)\}$$

Definition of persistence diagram

Definition

For a filtration

$$\mathbb{X} : X_1 \subset X_2 \subset \cdots \subset X_n ,$$

we compute homology groups with a field coefficient K

and obtain a sequence

$$H_q(\mathbb{X}) : H_q(X_1) \to H_q(X_2) \to \cdots \to H_q(X_n) .$$

This sequence is called a persistent homology.

In this talk, we set a filtration by the union of balls

$$X_r = \bigcup_{x_i \in X} B(x_i; r)$$

A persistent homology $H_q(\mathbb{X})$ can be seen as a representation of $\mathbf{A}_n$ -quiver.

Definition of persistence diagram

From Gabriel and Krull-Remak-Schmidt theorem, there is the following decomposition:

$$H_q(\mathbb{X}) \cong \bigoplus_{i \in I} \mathbb{I}[b_i, d_i] \ (I \text{ is a finite set})$$

$$\mathbb{I}[b,d] : 0 \longleftrightarrow 0 \longleftrightarrow \cdots \longleftrightarrow 0 \longleftrightarrow \overset{b}{\mathbb{F}} \longleftrightarrow \mathbb{F} \longleftrightarrow \cdots \longleftrightarrow \overset{d}{\mathbb{F}} \longleftrightarrow 0 \longleftrightarrow \cdots \longleftrightarrow 0$$

From the decomposition $H_q(\mathbb{X}) \cong \bigoplus_{i \in I} \mathbb{I}[b_i, d_i]$,

the persistence diagram is defined by $D_q(\mathbb{X}) = \{(b_i, d_i) \mid i \in I\}$.

Remark: $\bigoplus_r H_q(X_r)$ can be seen as a module over $K[t]$ and it can be decomposed from the structure theorem for PID.

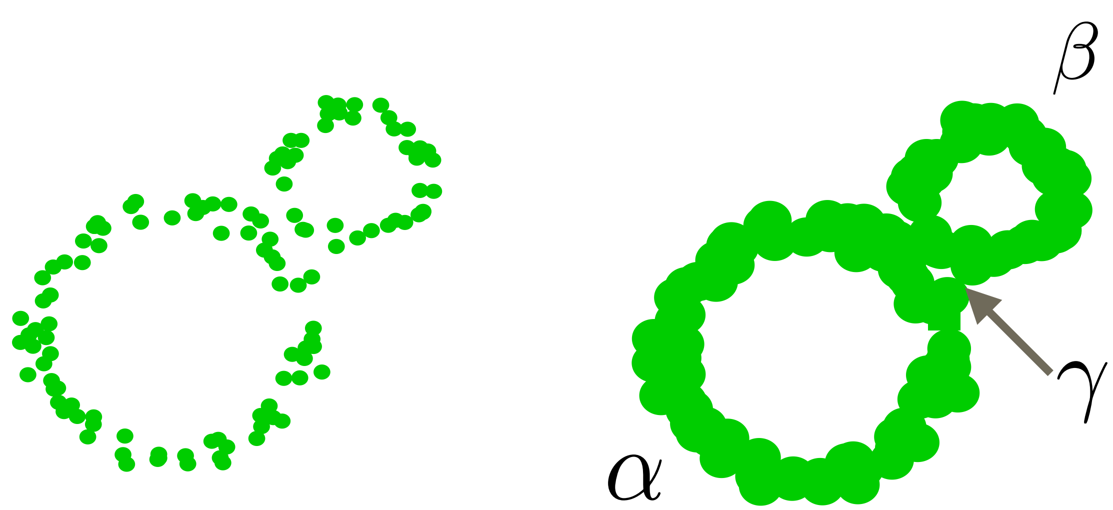
Persistence

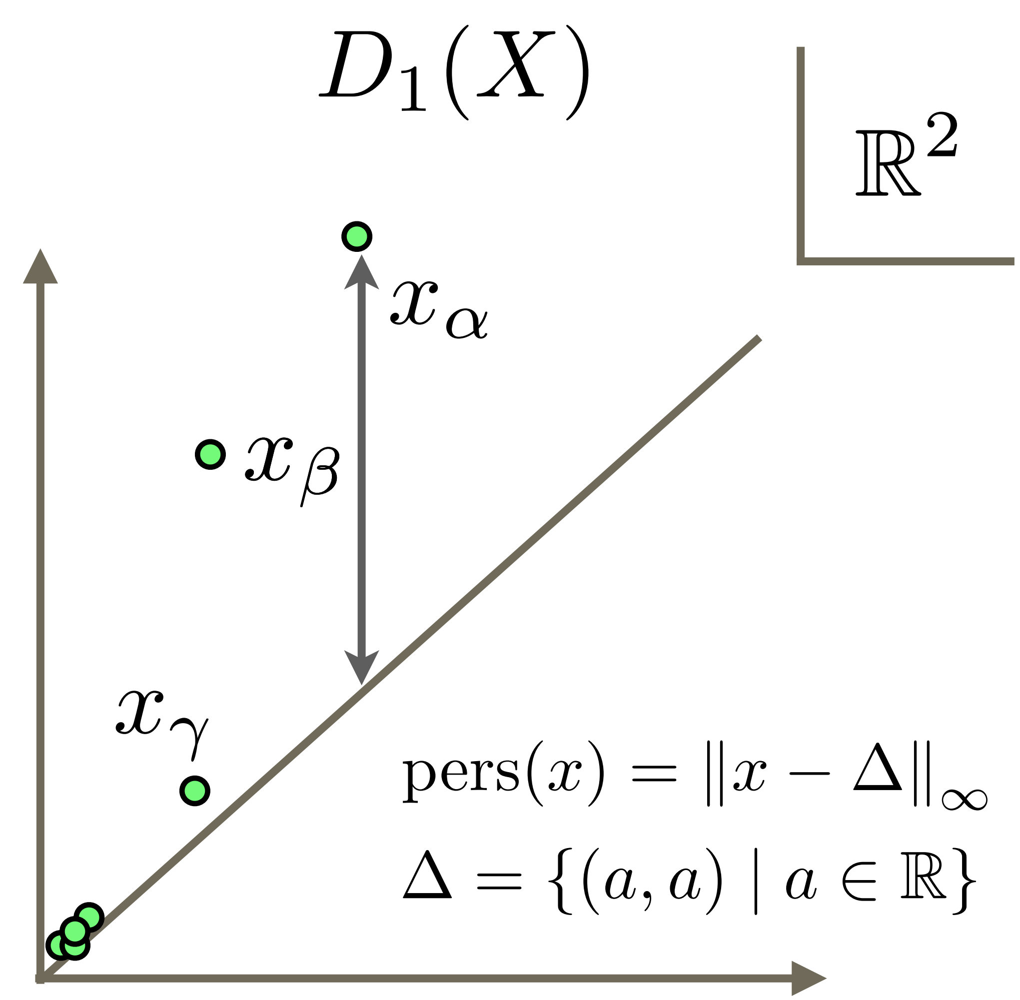
Definition

The lifetime of a cycle x_α

$$\mathrm{pers}(x_\alpha) = d_\alpha - b_\alpha$$

is called persistence.





A cycle with small persistence can be seen as a small cycle, and sometimes noisy cycle.

Metric structure of persistence diagram

The set of persistence diagrams is defined by

$$\mathcal{D} = \{D \mid D \text{ is a multiset in } \mathbb{R}^2_{ul} \text{ and } |D| < \infty\},$$

where $\mathbb{R}^2_{ul} = \{(b,d) \mid b \leq d \in \mathbb{R}\}$.

The **bottleneck (∞-Wasserstein) metric**

$$d_B(D,E) = \inf_{\gamma} \sup_{x \in D \cup \Delta} \|x - \gamma(x)\|_\infty \ (\gamma : D \cup \Delta \to E \cup \Delta \text{ is bijective}),$$

where $\Delta = \{(a,a) \mid a \in \mathbb{R}\}$ is the diagonal set,
becomes a distance on the set of persistence diagrams.

Remark: $(\mathcal{D}, d_B)$ is a metric space.

Stability theorem

Theorem[Cohen-Steiner et al., 2007]

For finite subsets $X, Y \subset \mathbb{R}^d$, $d_B(D_q(X), D_q(Y)) \leq d_H(X, Y)$,

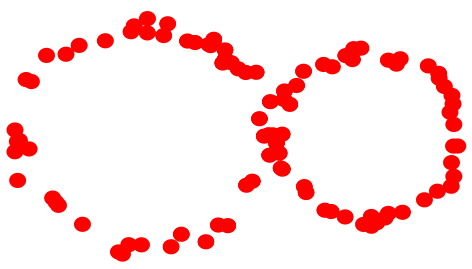
where $d_H(X,Y) = \max\left\{\max_{p\in X}\min_{q\in Y} d(p,q), \max_{q\in Y}\min_{p\in X} d(p,q)\right\}$ is the Hausdorff distance.

Significant property

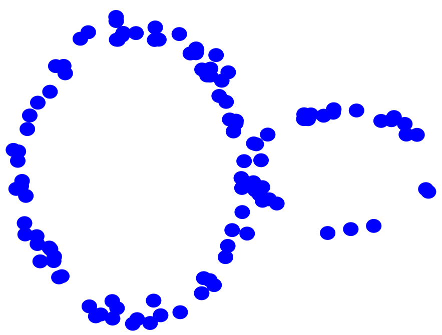
This map $X \rightarrow D_q(X)$ is Lipchitz continuous.

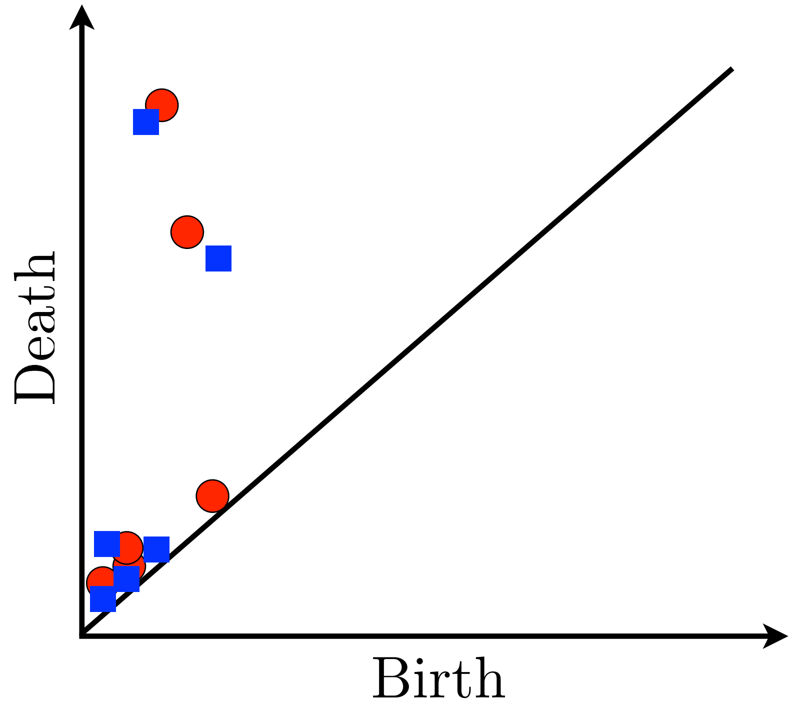
(Betti number $X \rightarrow \beta_q(X) = \dim H_q(X)$ is not continuous.)

X



Y





Stability theorem

Theorem[Cohen-Steiner et al., 2007]

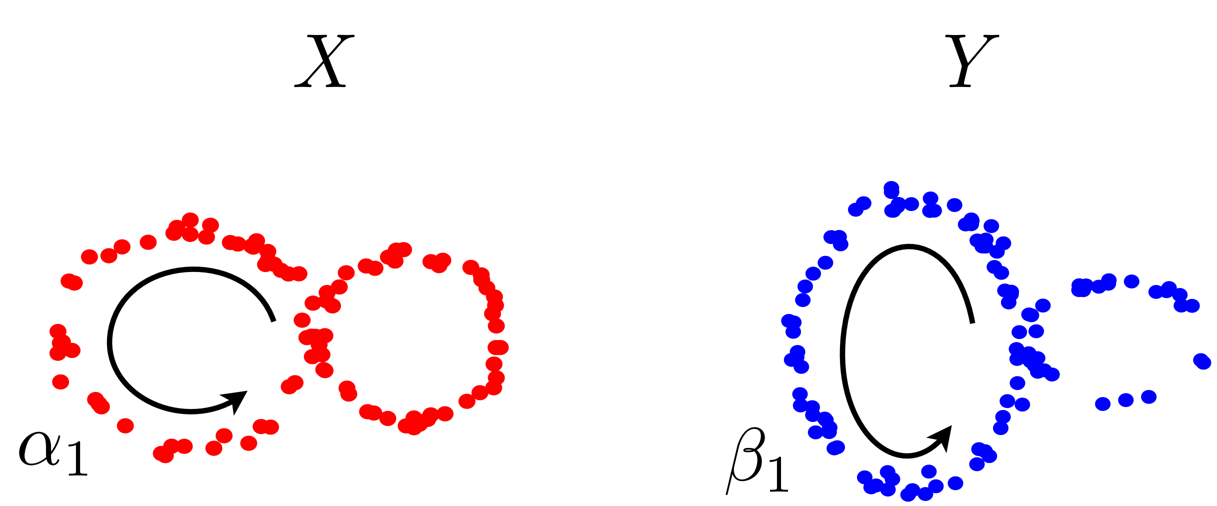
For finite subsets $X, Y \subset \mathbb{R}^d$, $d_B(D_q(X), D_q(Y)) \leq d_H(X, Y)$,

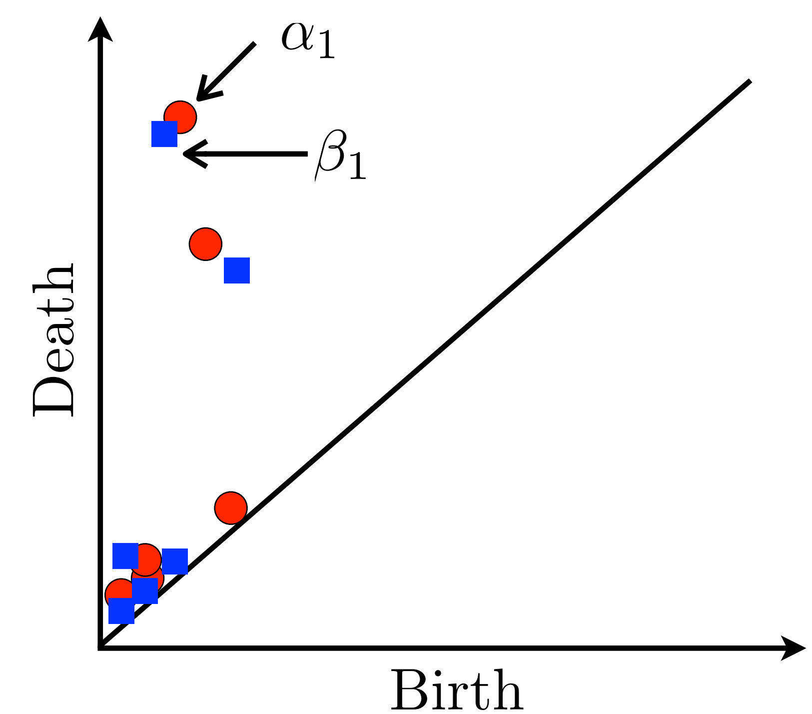
where $d_H(X,Y) = \max\left\{\max_{p \in X}\min_{q \in Y} d(p,q), \max_{q \in Y}\min_{p \in X} d(p,q)\right\}$ is the Hausdorff distance.

Significant property

This map $X \rightarrow D_q(X)$ is Lipchitz continuous.

(Betti number $X \rightarrow \beta_q(X) = \dim H_q(X)$ is not continuous.)





Statistical Topological Data Analysis

Data → Persistence diagram → Prediction / Classification / Testing / Estimation

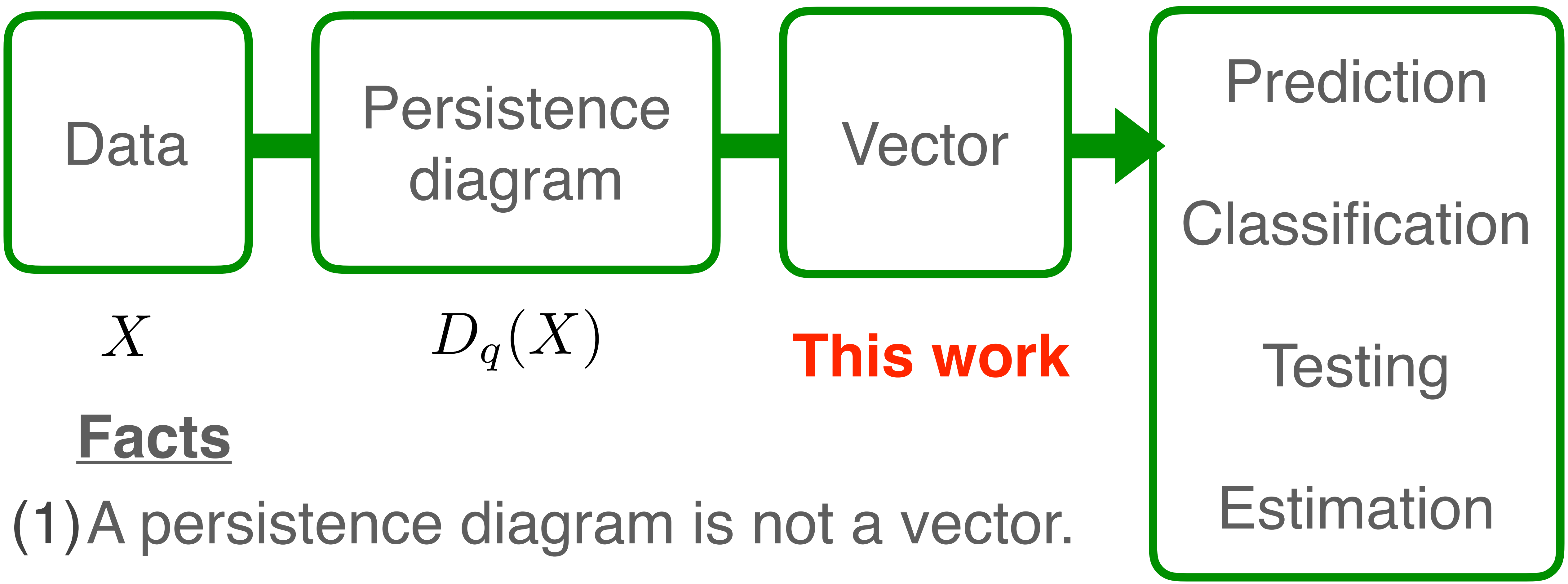
X $D_q(X)$

Facts

(1) A persistence diagram is not a vector.

(2) Standard statistical method is for vectors (multivariate analysis)

Statistical Topological Data Analysis



Facts

(1) A persistence diagram is not a vector.

(2) Standard statistical method is for vectors (multivariate analysis)

Make a vector representation of persistence diagram by kernel method

Section 2
Kernel method
~Statistical method for non-vector data~

Statistics for non-vector data

Let Ω be a data set and $x_1, \cdots, x_n \in \Omega$ be obtained data.

To consider statistical properties of the data, it is sometimes needed to calculate summaries, like mean/average:

$$x_1, \cdots, x_n \to \frac{1}{n}\sum_{i=1}^{n} x_i$$

To calculate statistical summaries, the data set Ω is desired to have structures of addition, multiplication by scalers, and inner product, that is, Ω should be an inner product space.

The space of persistence diagrams is not an inner space.

Statistics for non-vector data

While Ω does not always have an inner product, by defining a map $\phi : \Omega \to H$, where H is an inner product space, we can consider statistical summaries in H.

$$x_1, \cdots, x_n \to \phi(x_1), \cdots, \phi(x_n) \to \frac{1}{n}\sum_{i=1}^{n} \phi(x_i) \in H$$

(well-defined)

<u>Fact</u>

Many statistical summaries and machine learning techniques are calculated from the value of inner product:

$$\langle \phi(x_i), \phi(x_j) \rangle_H$$

Kernel method

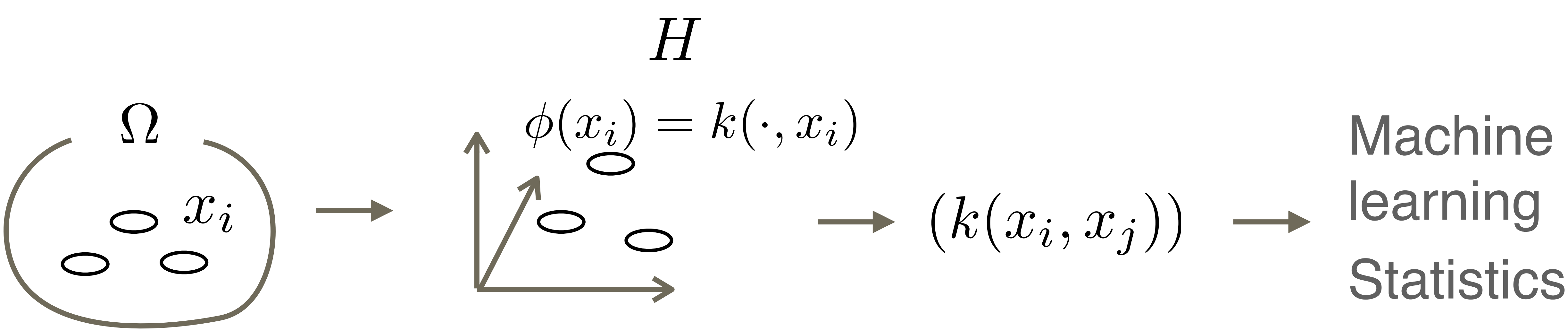
In kernel method, a positive definite kernel $k : \Omega \times \Omega \to \mathbb{R}$ is used as "non-linear" inner product on the data set.

$$k(x, y) = \langle \phi(x), \phi(y) \rangle_H$$

For an element $x \in \Omega$, $k(\cdot, x) : \Omega \to \mathbb{R}$ is a function and a vector in the functional space $C(\Omega)$.

In many cases, what we need is just the Gram matrix

$$(k(x_i, x_j))_{i,j=1,\cdots,n}$$



Kernel method

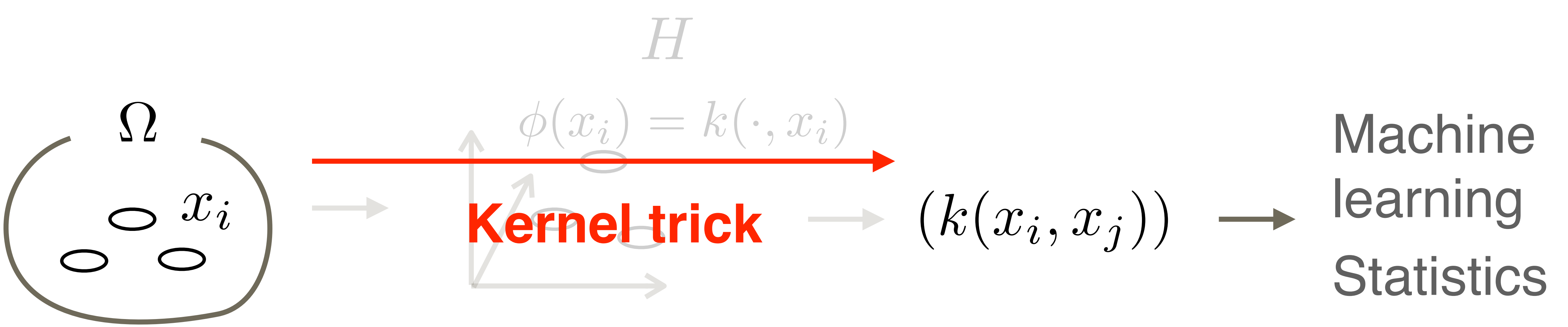
In kernel method, a positive definite kernel $k : \Omega \times \Omega \to \mathbb{R}$ is used as "non-linear" inner product on the data set.

$$k(x, y) = \langle \phi(x), \phi(y) \rangle_H$$

For an element $x \in \Omega$, $k(\cdot, x) : \Omega \to \mathbb{R}$ is a function and a vector in the functional space $C(\Omega)$.

In many cases, what we need is just the Gram matrix

$$(k(x_i, x_j))_{i,j=1,\cdots,n}$$



Kernel method

Definition

Let Ω be a set. A function $k : \Omega \times \Omega \to \mathbb{R}$ is called a positive definite kernel when k satisfies

- $k(x, y) = k(y, x)$
- For any $x_1, \ldots, x_n \in \Omega$, the matrix $(k(x_i, x_j))$ is semi-positive definite (this matrix is called the Gram matrix).

Examples in case of $\Omega = \mathbb{R}^d$,

- linear kernel $k_L(x, y) = \langle x, y \rangle$
- polynominal kernel $k_P(x, y) = (\langle x, y \rangle + c)^d$
- **Gaussian kernel** $k_G(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$

Kernel method

Moore-Aronszajn theorem

A p.d. kernel k uniquely defines a Hilbert space $\mathcal{H}_k$ which is called reproducing kernel Hilbert space (RKHS) satisfying

- for any $x \in \Omega$, the function $k(\cdot, x) : \Omega \to \mathbb{R}$ is in $\mathcal{H}_k$.
- $\mathrm{Span}\{k(\cdot, x) \mid x \in \Omega\}$ is dense in $\mathcal{H}_k$.
- for any $x \in \Omega$ and $f \in \mathcal{H}_k$, $\langle f, k(\cdot, x)\rangle_{\mathcal{H}_k} = f(x)$.

That is, $k(\cdot, x) : \Omega \to \mathbb{R}$ is an element of the Hilbert space $\mathcal{H}_k$.
(**RKHS vector**)

The inner product is computed by

$$\langle k(\cdot, y), k(\cdot, x)\rangle_{\mathcal{H}_k} = k(x, y)\ .$$

Kernel embedding

In order to consider statistical properties of (probability) distributions on Ω, we vectorize them by a positive definite kernel.

Here, let Ω be a locally compact Hausdorff space and $M_b(\Omega)$ be the set of finite signed Radon measures. Then, measures can be represented as an element of RKHS:

$$M_b(\Omega) \ni \mu \mapsto E_k(\mu) := \int k(\cdot, x) d\mu(x) \in \mathcal{H}_k$$

This map is called kernel embedding and this integral is interpreted as the Bochner integral.

Kernel embedding

If a kernel is "nice", the kernel embedding becomes injective.

$$M_b(\Omega) \ni \mu \mapsto E_k(\mu) := \int k(\cdot, x) d\mu(x) \in \mathcal{H}_k$$

The Gaussian kernel $k_G(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$ is C_0-universal.

Definition

A p.d. kernel k is said to be C_0**-universal** if $k(\cdot, x)$ is in $C_0(\Omega)$ and $\mathcal{H}_k$ is dense in $C_0(\Omega)$.

Proposition [B.K. Sriperumbudur et al., 2011]

If k is C_0-universal, E_k is injective, and hence

$d_k(\mu, \nu) := \|E_k(\mu) - E_k(\nu)\|_{\mathcal{H}_k}$ becomes a metric on $M_b(\Omega)$.

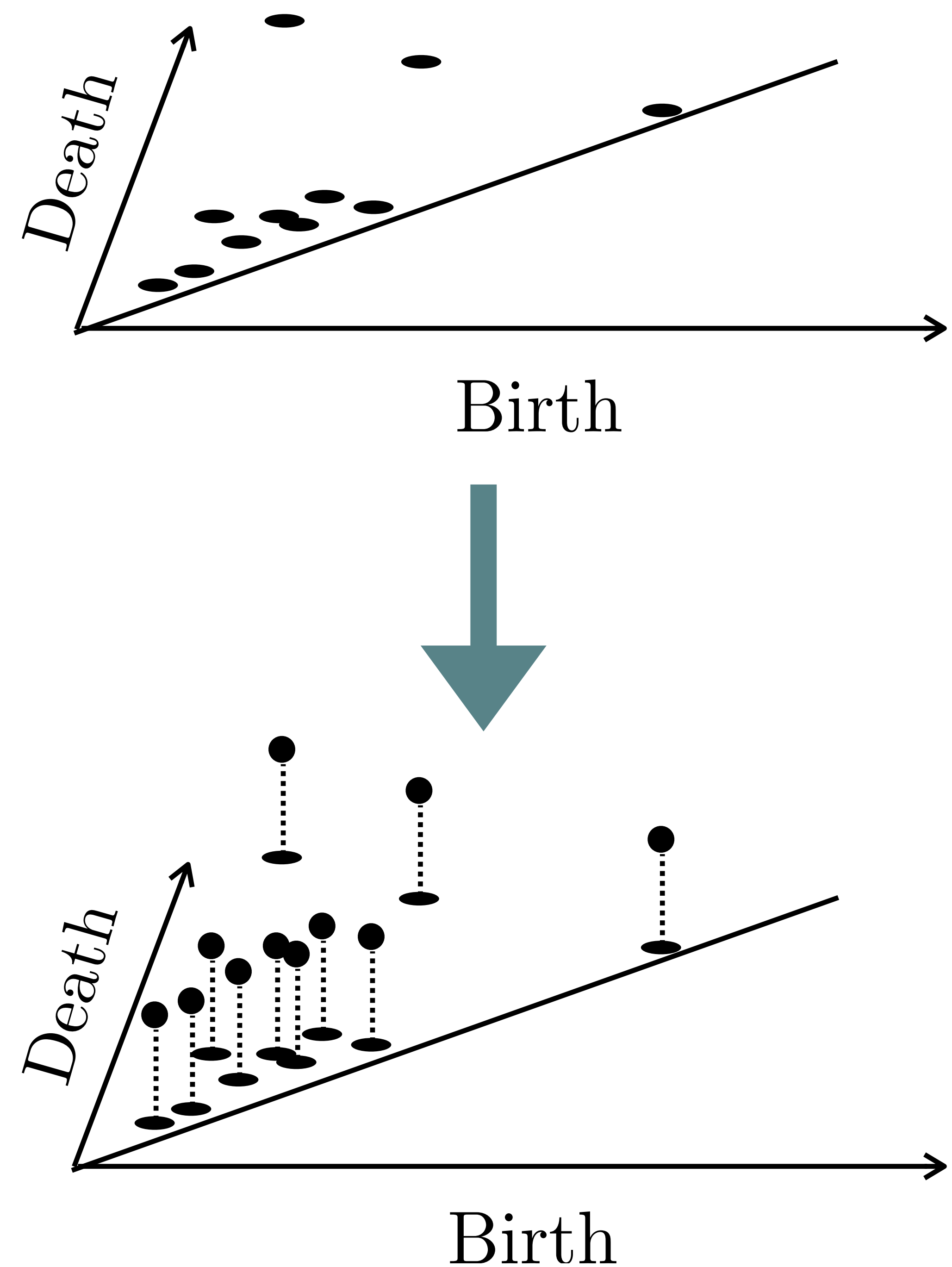
Section 3
Kernel on persistence diagram

Kernel on persistence diagram

A persistence diagram can be seen as a counting measure

$$\mu_D = \sum_{x \in D} \delta_x$$

where δ_x is the Delta measure.



<u>Observation</u>
Point close to the diagonal has a small persistence, so it sometimes can be seen as a noisy cycle.

Kernel on persistence diagram

A persistence diagram can be seen as a counting measure

$$\mu_D = \sum_{x \in D} \delta_x$$

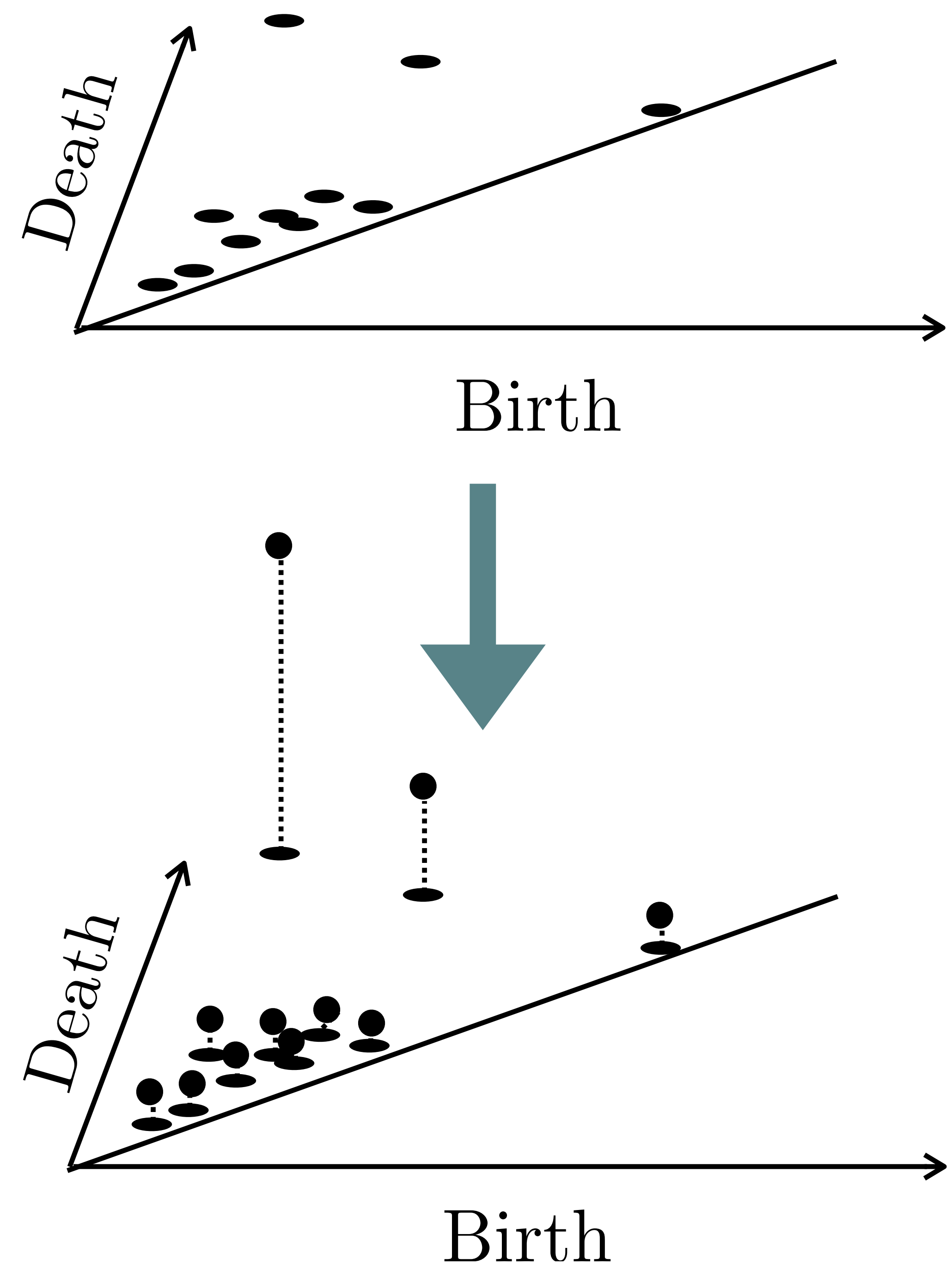
where δ_x is the Delta measure.

By defining an appropriate function

$$w : \mathbb{R}^2 \to \mathbb{R} \ ,$$

a persistence diagram is represented as a **weighted** measure:

$$\mu_D^w = \sum_{x \in D} w(x)\delta_x$$



Kernel on persistence diagram

For the weighted measure, we consider the kernel embedding:

$$D \mapsto \mu_D^w \mapsto E_k(\mu_D^w) = \sum_{x \in D} w(x) k(\cdot, x) \in \mathcal{H}_k$$

Then, we define a kernel on persistence diagram as the Gaussian kernel on the RKHS:

$$K_G(D, E) = \exp\left(-\frac{1}{2\tau^2} \left\| E_k(\mu_D^w) - E_k(\mu_E^w) \right\|_{\mathcal{H}_k}^2 \right)$$

We can define a linear kernel on RKHS

$$K_L(D, E) = \langle E_k(\mu_D^w), E_k(\mu_E^w) \rangle_{\mathcal{H}_k}$$

Metric on RKHS vectors

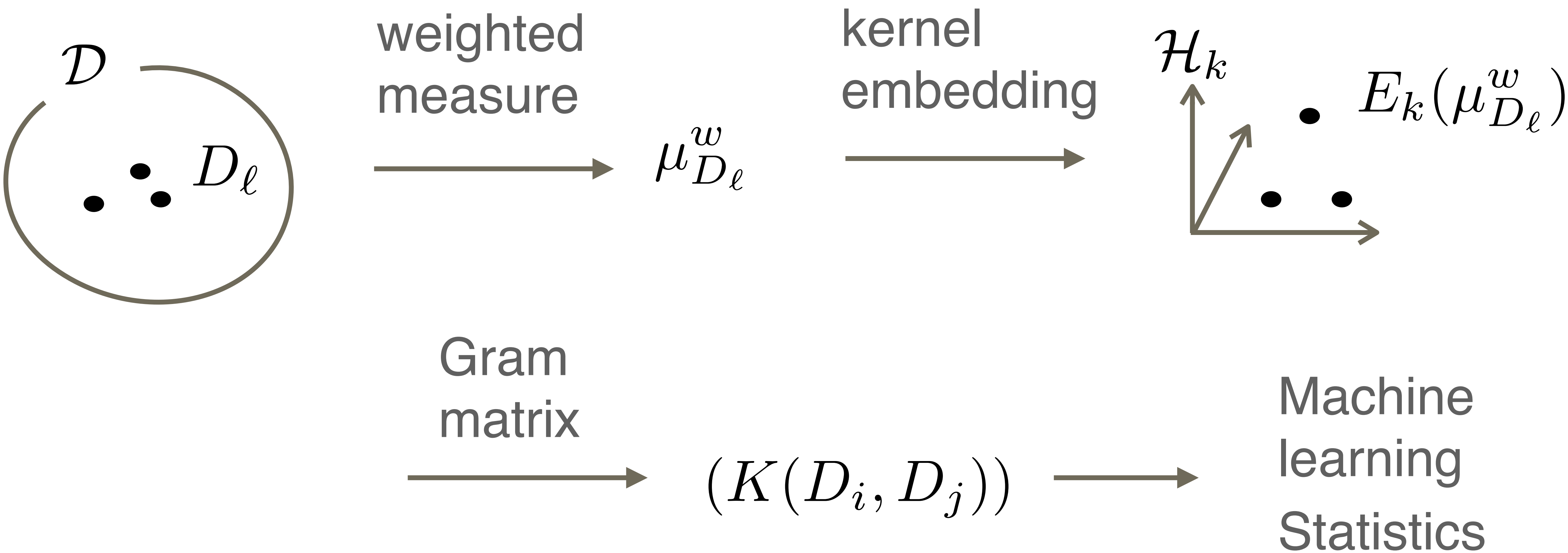
While the vector $E_k(\mu_D^w)$ is an element of $\mathcal{H}_k \subset C(\mathbb{R}^2_{ul})$, the inner product is easy to compute:

$$\langle E_k(\mu_D^w), E_k(\mu_E^w)\rangle_{\mathcal{H}_k} = \sum_{x\in D}\sum_{y\in E} w(x)w(y)k(x,y)\,.$$

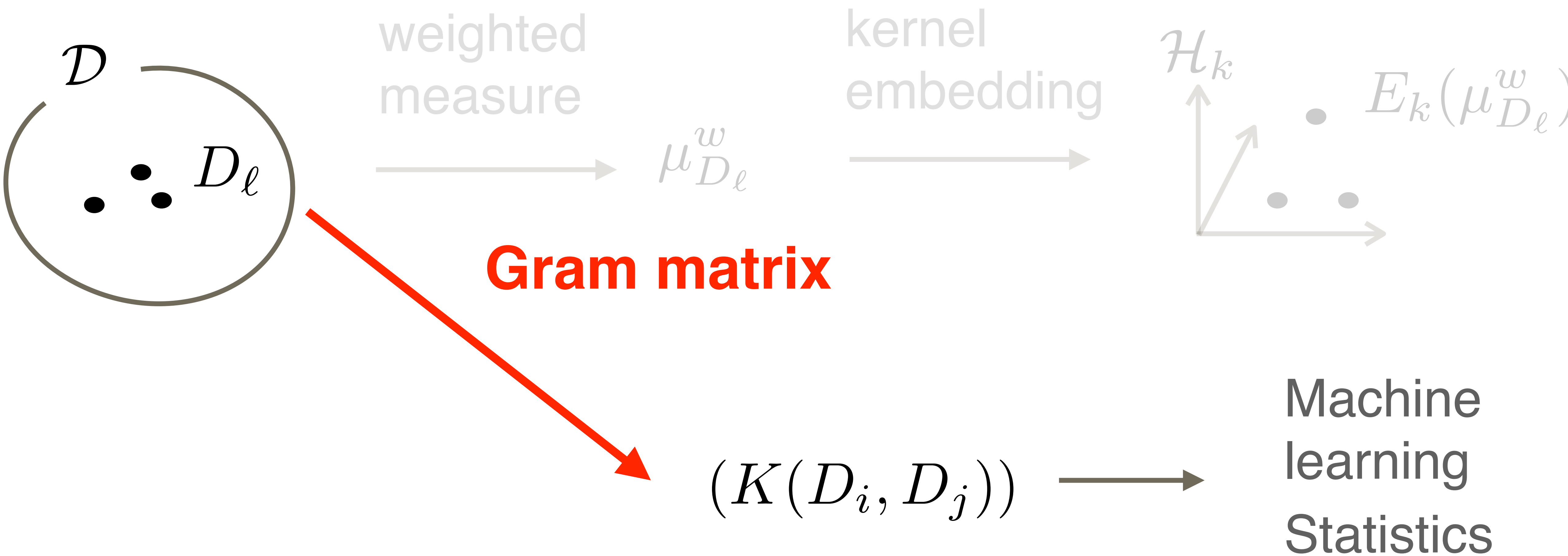
The distance also has the following expansion:

$$\begin{aligned}\|E_k(\mu_D^w) - E_k(\mu_E^w)\|^2_{\mathcal{H}_k} = &\sum_{x\in D}\sum_{x'\in D} w(x)w(x')k(x,x') + \sum_{y\in E}\sum_{y'\in E} w(y)w(y')k(y,y') \\ &- 2\sum_{x\in D}\sum_{y\in E} w(x)w(y)k(x,y)\end{aligned}$$

Kernel on persistence diagram



Kernel on persistence diagram



$$K_G(D,E) = \exp\left(-\frac{1}{2\tau^2}\|E_k(\mu^w_D) - E_k(\mu^w_E)\|^2_{\mathcal{H}_k}\right)$$

$$\|E_k(\mu^w_D) - E_k(\mu^w_E)\|^2_{\mathcal{H}_k} = \sum_{x\in D}\sum_{x'\in D} w(x)w(x')k(x,x') + \sum_{y\in E}\sum_{y'\in E} w(y)w(y')k(y,y') - 2\sum_{x\in D}\sum_{y\in E} w(x)w(y)k(x,y)$$

computable

Choice of functions in PWGK

$$\begin{array}{ccc} \mathcal{D} & \longrightarrow & \mathcal{H}_k \\ \Downarrow & & \Downarrow \\ D & \longmapsto & E_k(\mu_D^w) = \sum_{x \in D} w(x) k(\cdot, x) \end{array}$$

To obtain a RKHS representation, we have used a weight function w and a kernel k. In this talk, we set the following functions:

$$k_G(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$$

$$w_{\mathrm{arc}}(x) = \arctan(C\mathrm{pers}(x)^p) \ (C, p > 0)$$

$$\mathrm{pers}(x) = d - b \ (x = (b, d) \in \mathbb{R}^2_{ul})$$

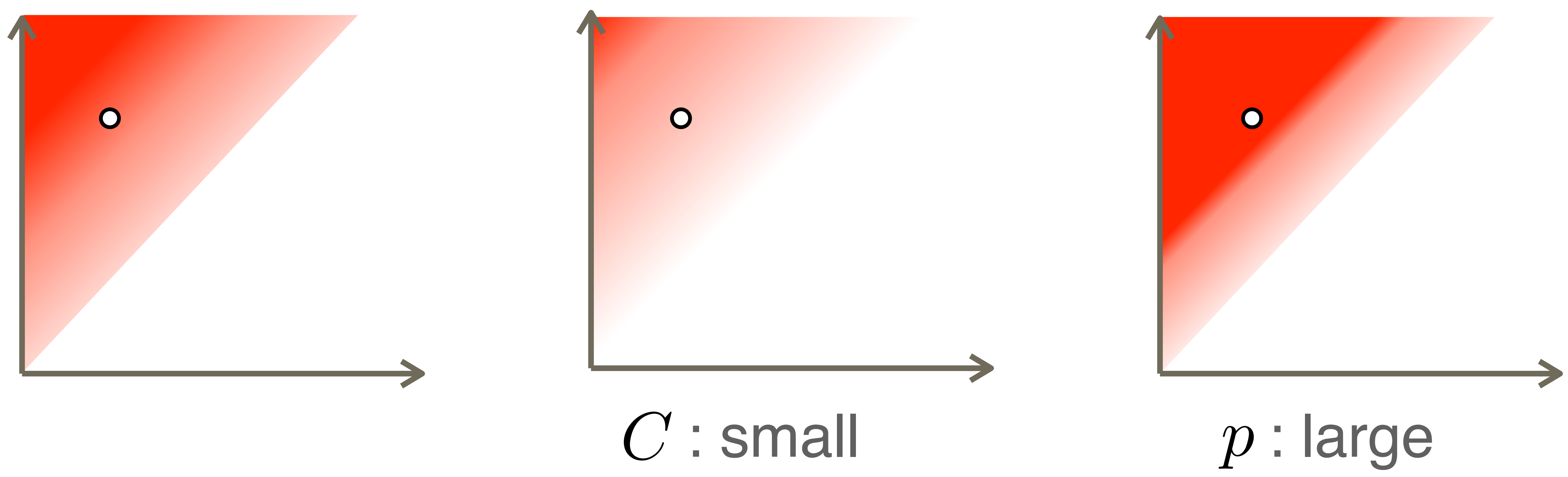
The reason of w_{arc} is for stability result.

Weight function

Parameters appearing $w_{\mathrm{arc}}(x) = \arctan(C \mathrm{pers}(x)^p)$ control

C : where is a noisy region,

p : how sharp is the noisy decision boundary.



Strong (light) red mean that the value of weight function is high (low).

Stability theorem

Main theorem [K, Fukumizu, Hiraoka, 2016]

For a compact subset $M \subset \mathbb{R}^d$ and two finite point sets X, Y in M, if $p > d+1$, then

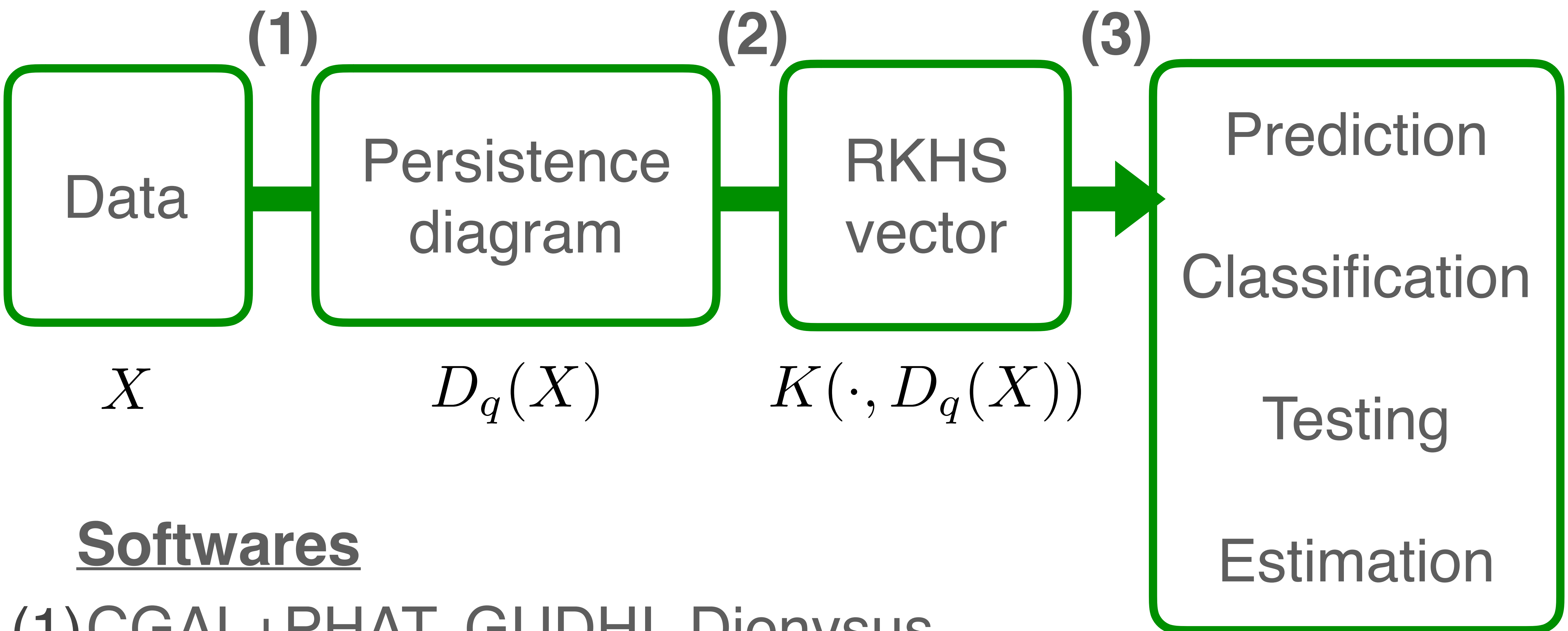
$$\left\| E_{k_G}(\mu^{w_{\mathrm{arc}}}_{D_q(X)}) - E_{k_G}(\mu^{w_{\mathrm{arc}}}_{D_q(Y)}) \right\|_{\mathcal{H}_{k_G}} \leq L(M,d;C,p,\sigma) d_B(D_q(X), D_q(Y)) \; ,$$

where

$$L(M,d;C,p,\sigma) = \left\{ \frac{\sqrt{2}}{\sigma} \frac{p}{p-d} C_M \mathrm{diam}(M)^{p-d} + \frac{4p(p-1)}{p-1-d} C_M \mathrm{diam}(M)^{p-1-d} \right\} C \; .$$

- Kernel embedding $D \mapsto \sum_{x \in D} w(x) k(\cdot, x)$ is **injective** and **continuous** (k_G is a C_0-universal kernel)

Statistical Topological Data Analysis



Softwares

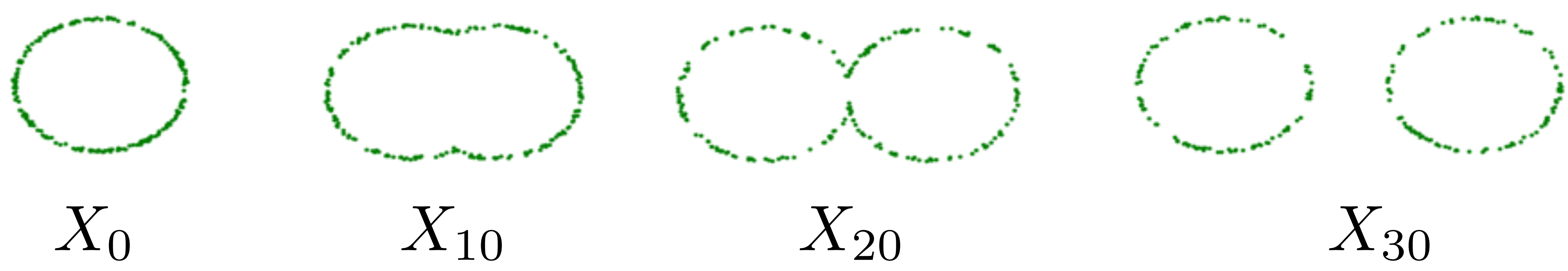
(1) CGAL+PHAT, GUDHI, Dionysus, R-TDA, Ripser, **HomCloud**(Hiraoka lab), …

(2) Kernel on persistence diagrams (This work)

(3) libsvm, e-1071, scikit-learn, …

Section 4
Demonstrations

Synthesized two circles data



$$d = 0.05k \ (k = 0, \cdots, 40)$$

$$\begin{cases} x_1(i) &= \cos(t_1^i) - d \\ y_1(i) &= \sin(t_1^i) \end{cases} \qquad t_1^i \sim \begin{cases} [\arccos(d), 2\pi - \arccos(d)] \ (d < 1) \\ [0, 2\pi] \ (d > 1) \end{cases}$$

$$\begin{cases} x_2(i) &= \cos(t_2^i) + d \\ y_2(i) &= \sin(t_2^i) \end{cases} \qquad t_2^i \sim \begin{cases} [-\pi + \arccos(d), \pi - \arccos(d)] \ (d < 1) \\ [-\pi, \pi] \ (d > 1) \end{cases}$$

$$X_k := \{(x_1(i), y_1(i))\}_{i=1,\cdots,100} \cup \{(x_2(i), y_2(i))\}_{i=1,\cdots,100}$$

Goal : Detect the change point in this system

Procedures

1. **Prepare data (two circles data)**

$$X_1, \cdots, X_n$$

2. **Compute persistence diagrams**

$$D(X_1), \cdots, D(X_n)$$

3. **Compute the Gram matrix**

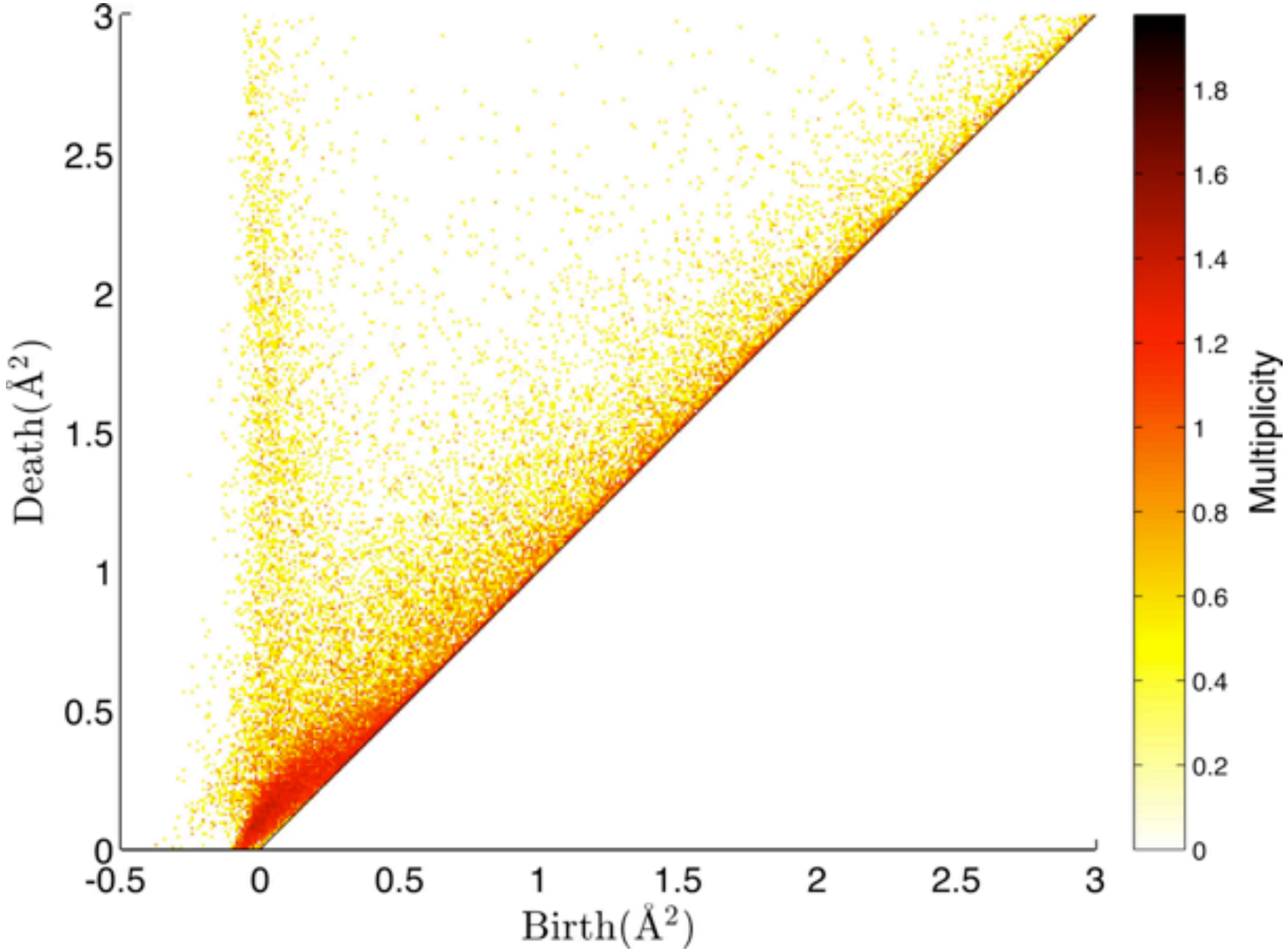
$$K_G(D(X_i), D(X_j)) = \exp\left(-\frac{1}{2\tau^2}\left\|E_k(\mu^w_{D(X_i)}) - E_k(\mu^w_{D(X_j)})\right\|^2_{\mathcal{H}_k}\right)$$

4. **Apply statistical methods**
   kernel change point analysis
   kernel principal component analysis (kPCA)

Section 5
Applications

Application for glass transition problem

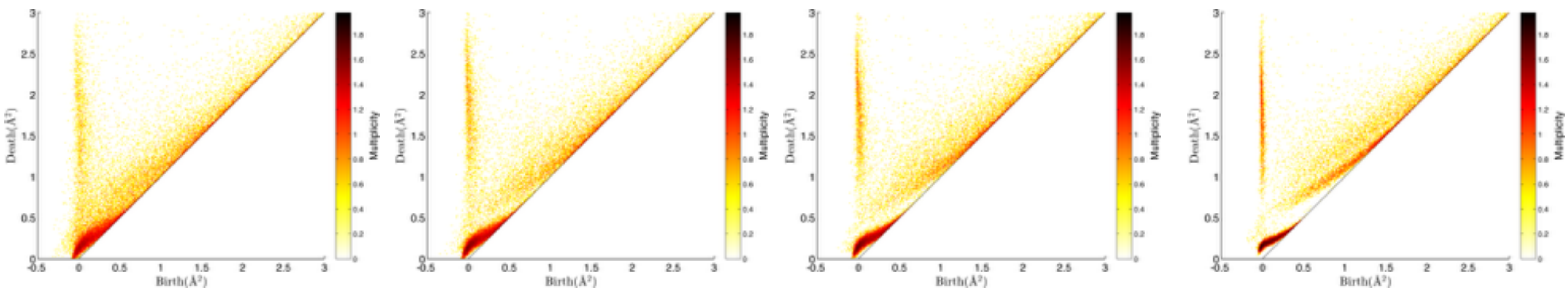
Data are atomic configurations of silica obtained from several temperatures (from liquid to glass).



Application for glass transition problem

Data are atomic configurations of silica obtained from several temperatures.

$D_1 \quad \cdots \quad D_\ell \qquad D_{\ell+1} \quad \cdots \quad D_{80}$

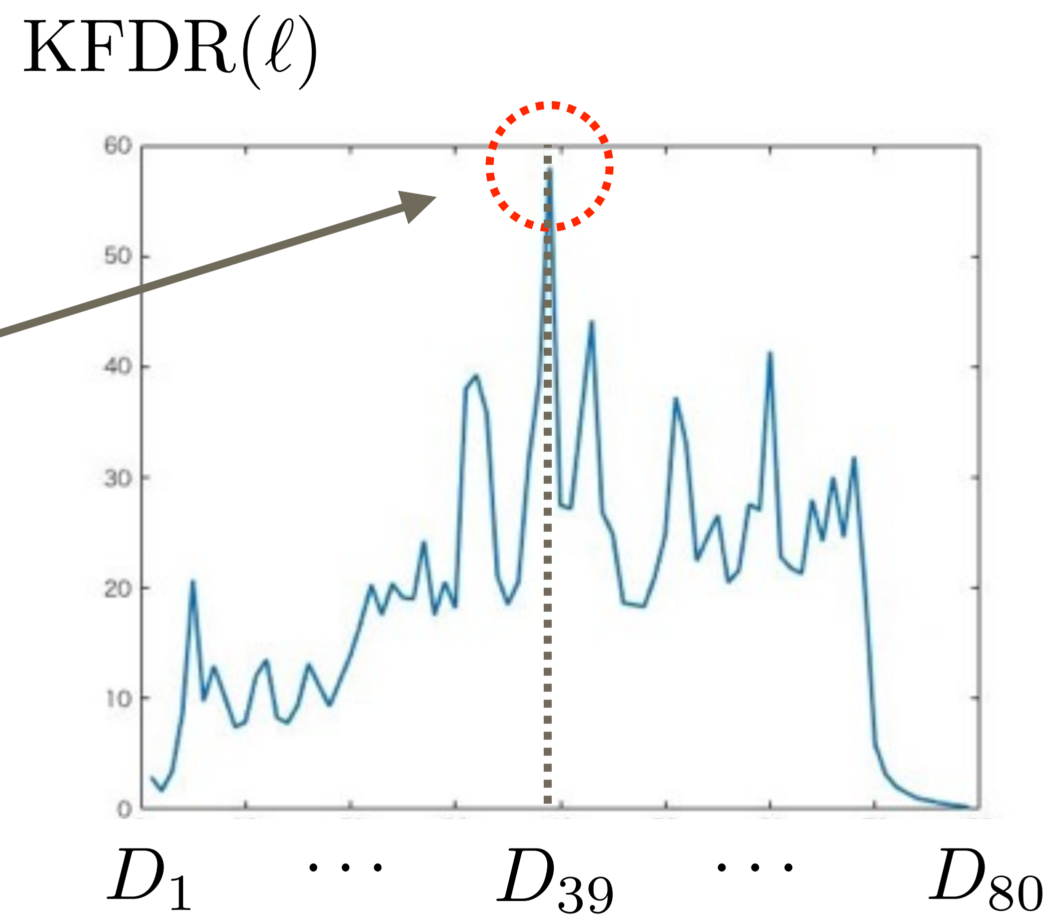


Liquid ⟷ Glass

Where (which index ℓ) is a transition point?

What we have to do is just to compute the Gram matrix of persistence diagrams $(K(D_i, D_j))_{i,j=1,\ldots,n}$ (Kernel method).

Application for glass transition problem



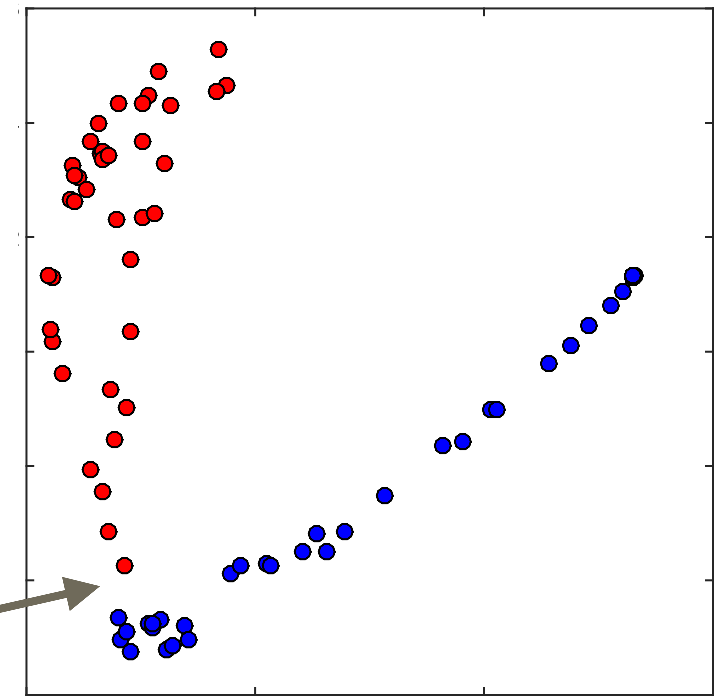
The change point is estimated from the peak of the KFDR graph.

This KFDR result tells us that $D_{39} \simeq 3100\mathrm{K}$ is an estimated change point and it is in the liquid-glass transition interval $[2000\mathrm{K}, 3500\mathrm{K}]$ which is determined by physics.

Remark: Our method uses only topological structure of silica.

Application for glass transition problem

3100K



Contribution rate : 81.9%

The 2-dimensional kernel principal component analysis (PCA) plot of persistence diagrams and colors are assigned before and after $3100\mathrm{K}$.

Conclusion

Our contribution : Kernel based statistical framework on persistence diagram

Acknowledgment : Takenobu Nakamura (Tohoku University)
Ippei Obayashi (Tohoku University)
Emerson Escolar (Tohoku University)

Reference

G. Kusano, Kenji Fukumizu, and Yasuaki Hiraoka. Persistence weighted Gaussian kernel for topological data analysis, ICML, pp. 2004–2013, 2016.

Thank you for your attention