Probability measures on the space of persistence diagrams

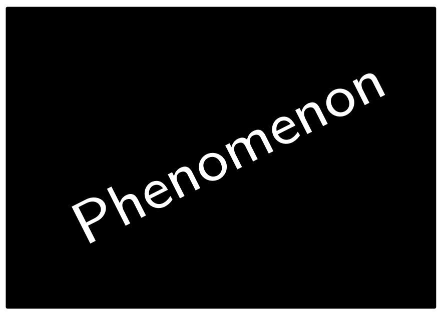
Yuriy Mileyko
yury@math.duke.edu

Joint work with
Sayan Mukherjee and John Harer

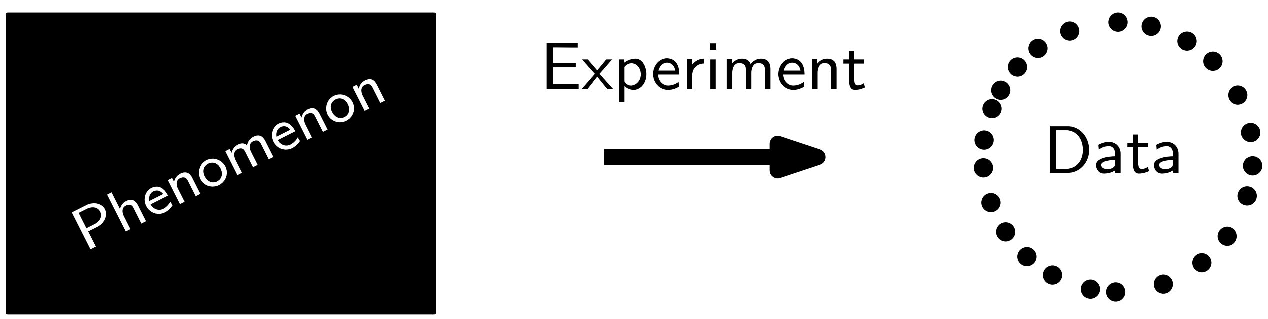
Workshop on Computational Topology
Fields Institute
November 8, 2011

General Problem

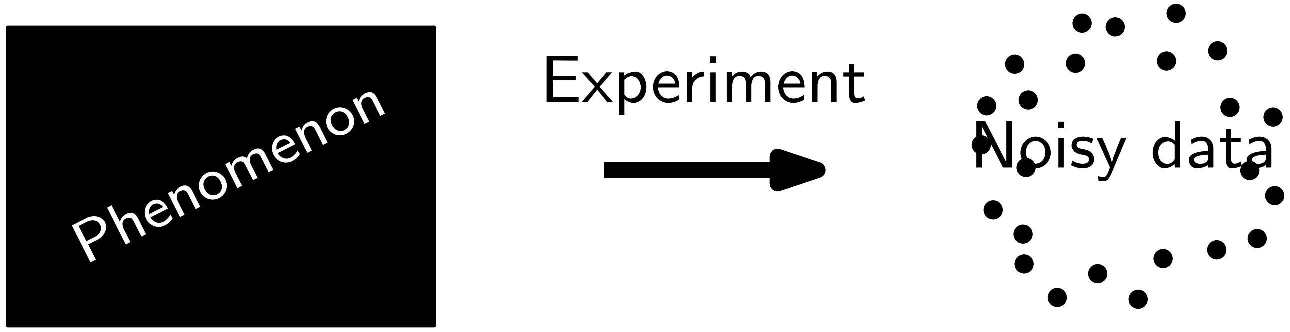
General Problem



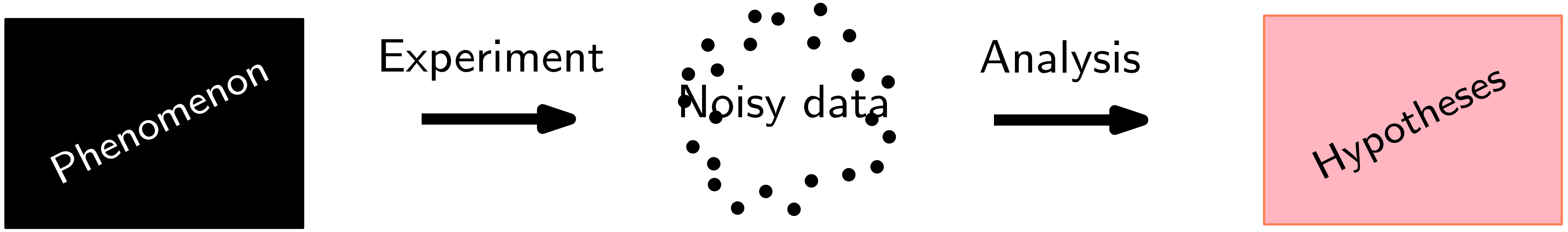
General Problem



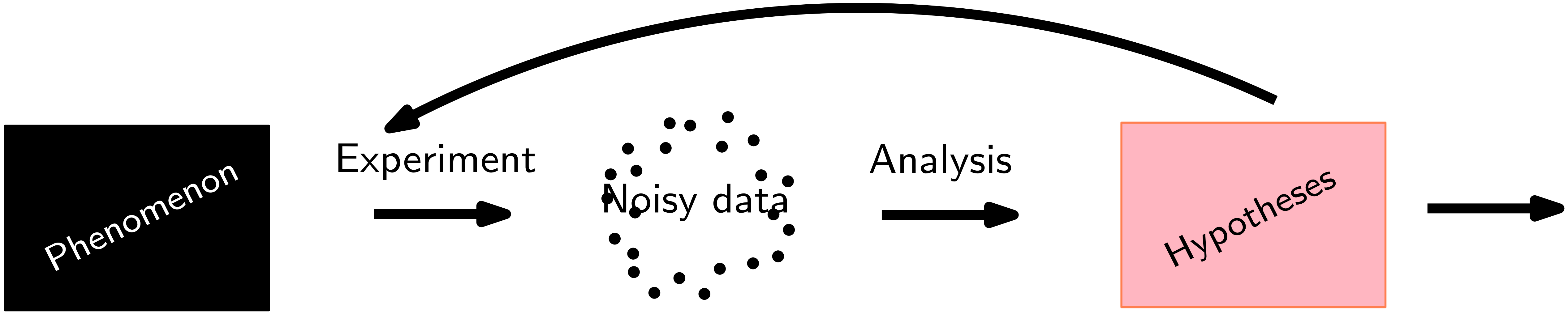
General Problem



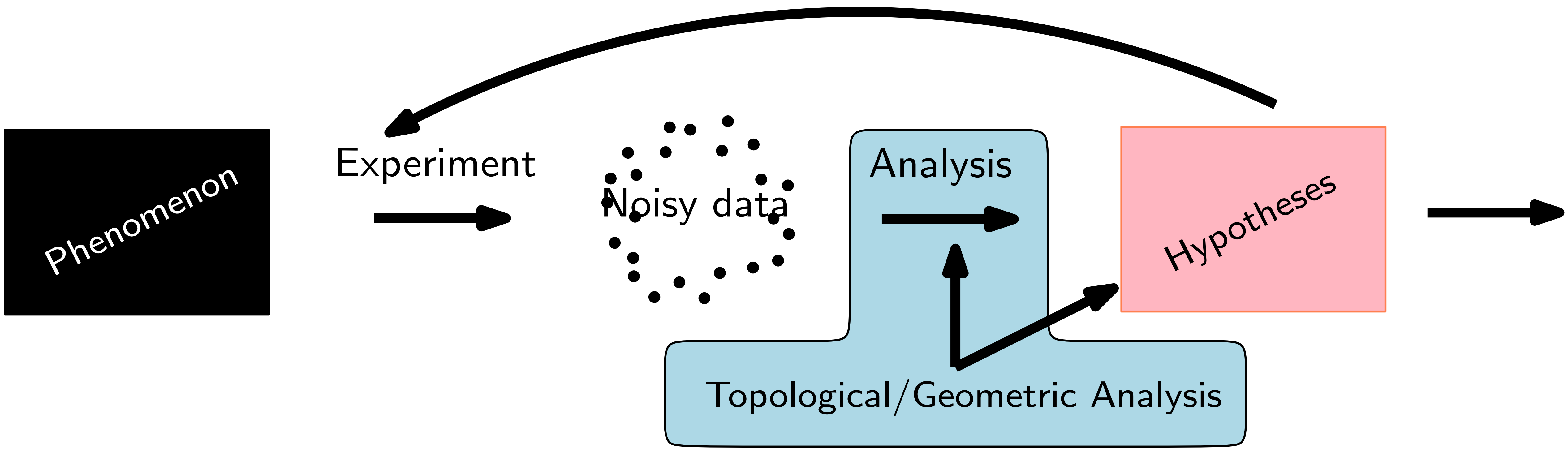
General Problem



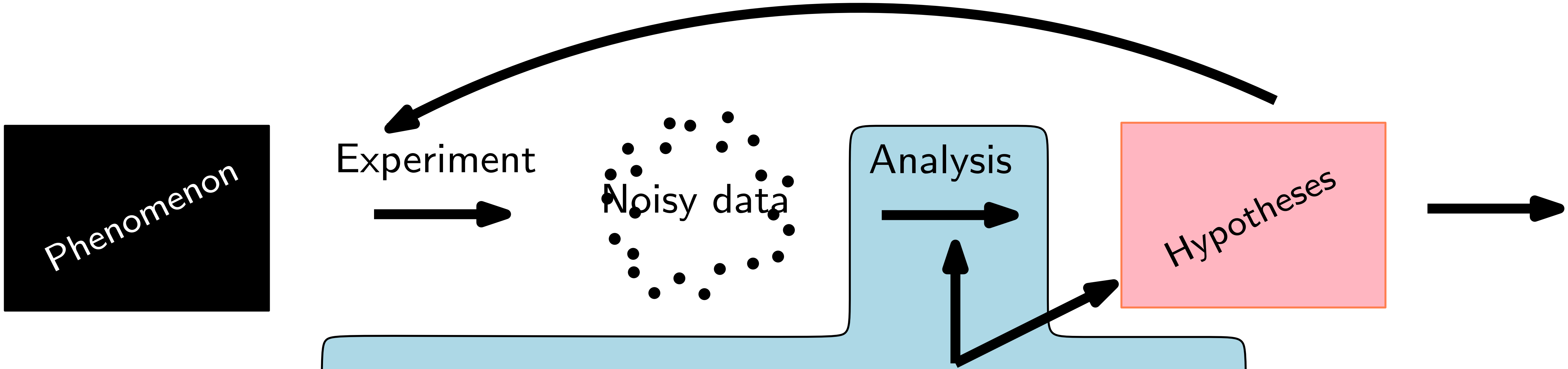
General Problem



General Problem

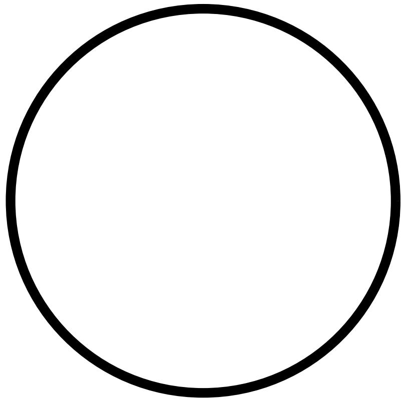


General Problem

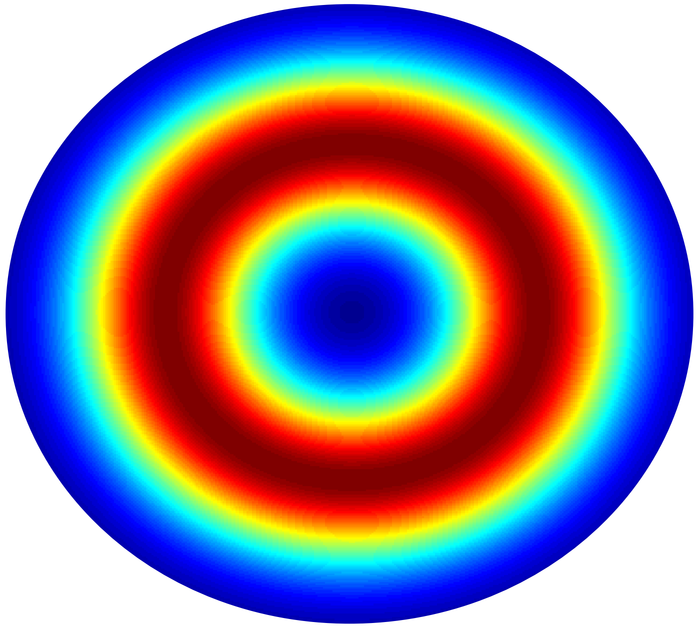


- Capture topological (and geometric) properties of data with topological summaries.
- Use topological summaries to perform a particular type of inference on the data.

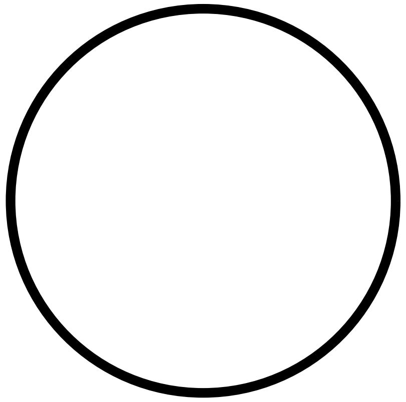
Specific Problem



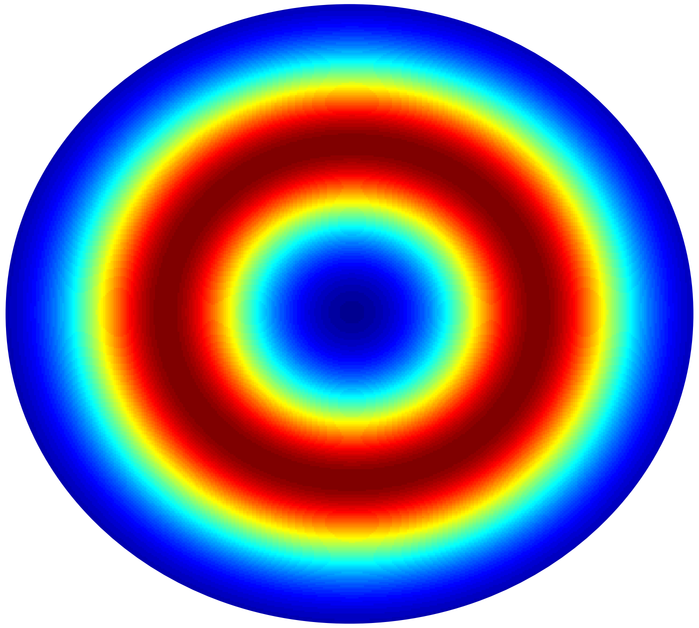
Suppose we sample a circle with Gaussian noise

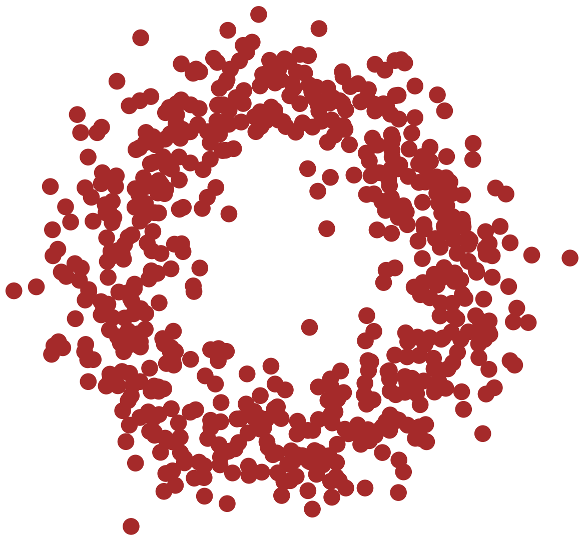


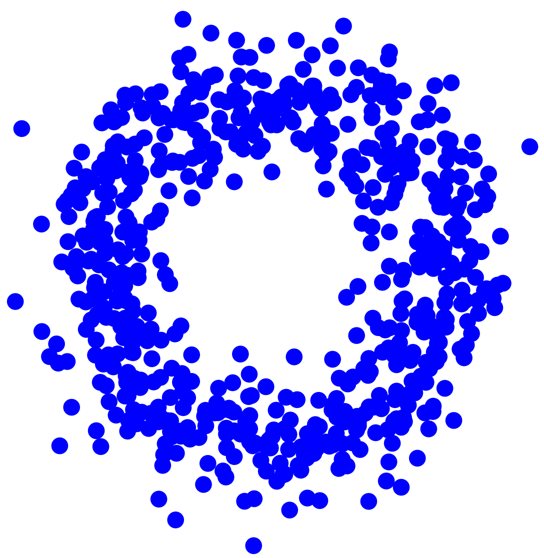
Specific Problem

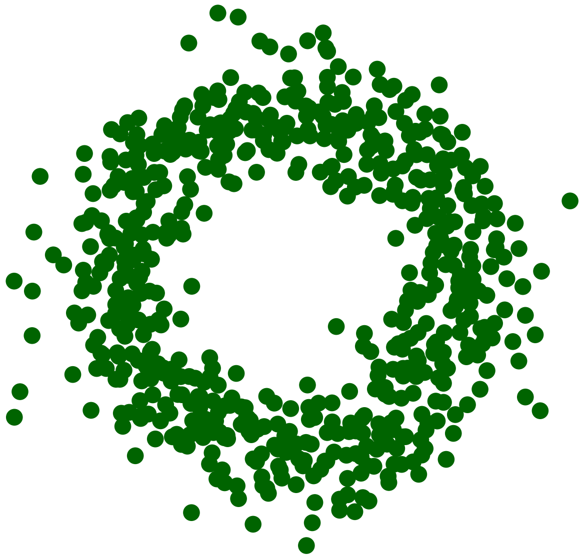


Suppose we sample a circle with
Gaussian noise

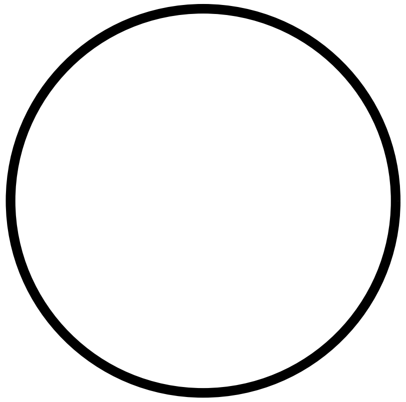




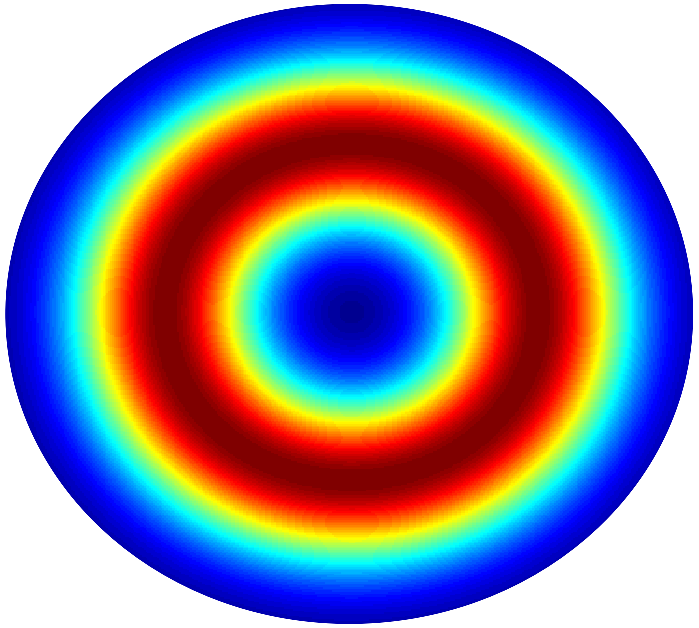


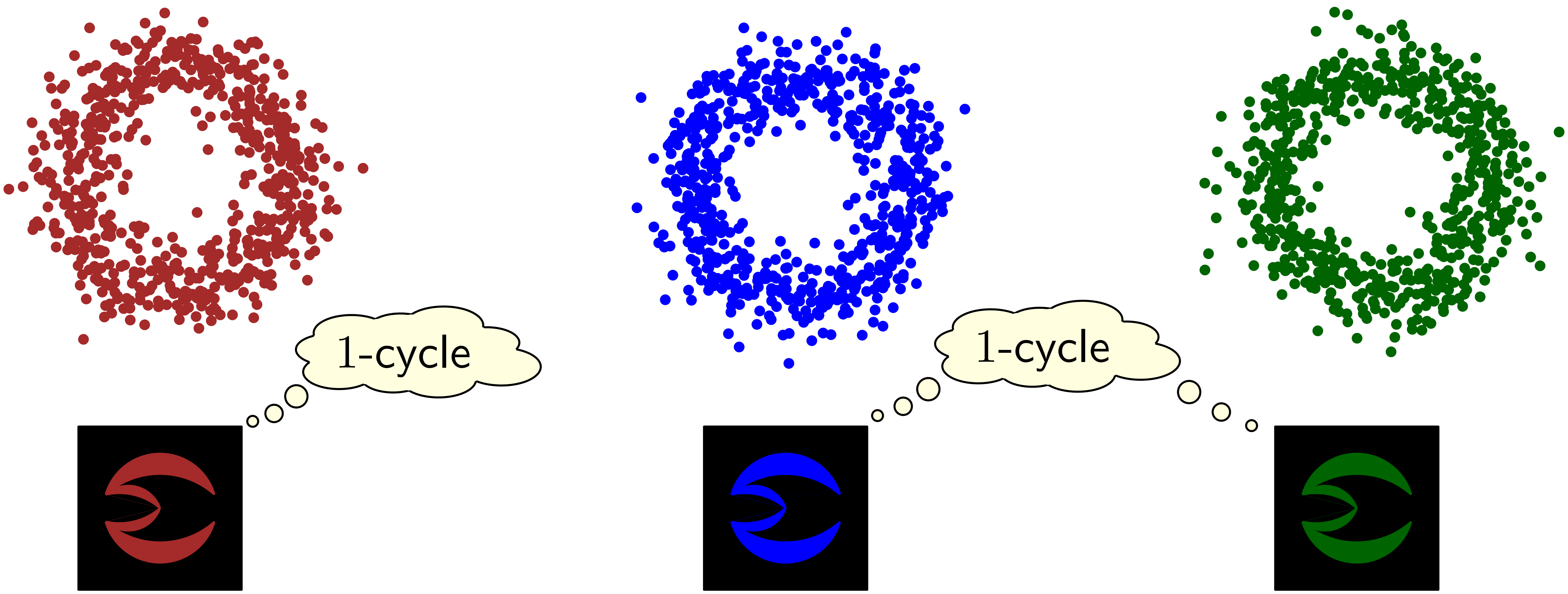


Specific Problem

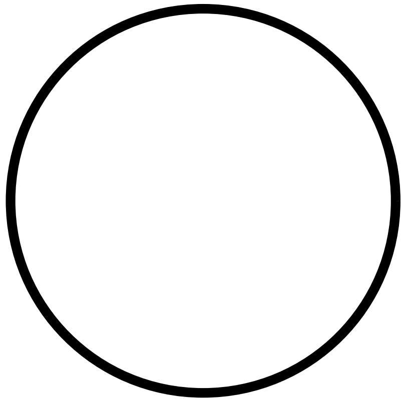


Suppose we sample a circle with Gaussian noise

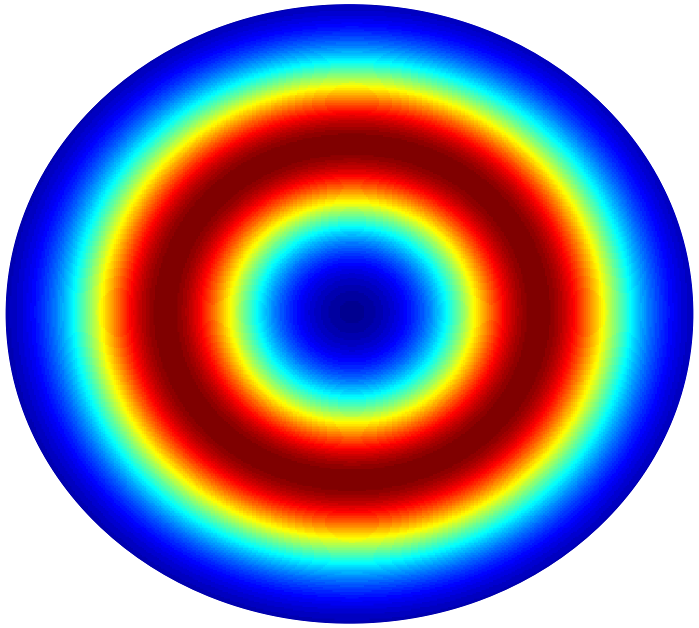


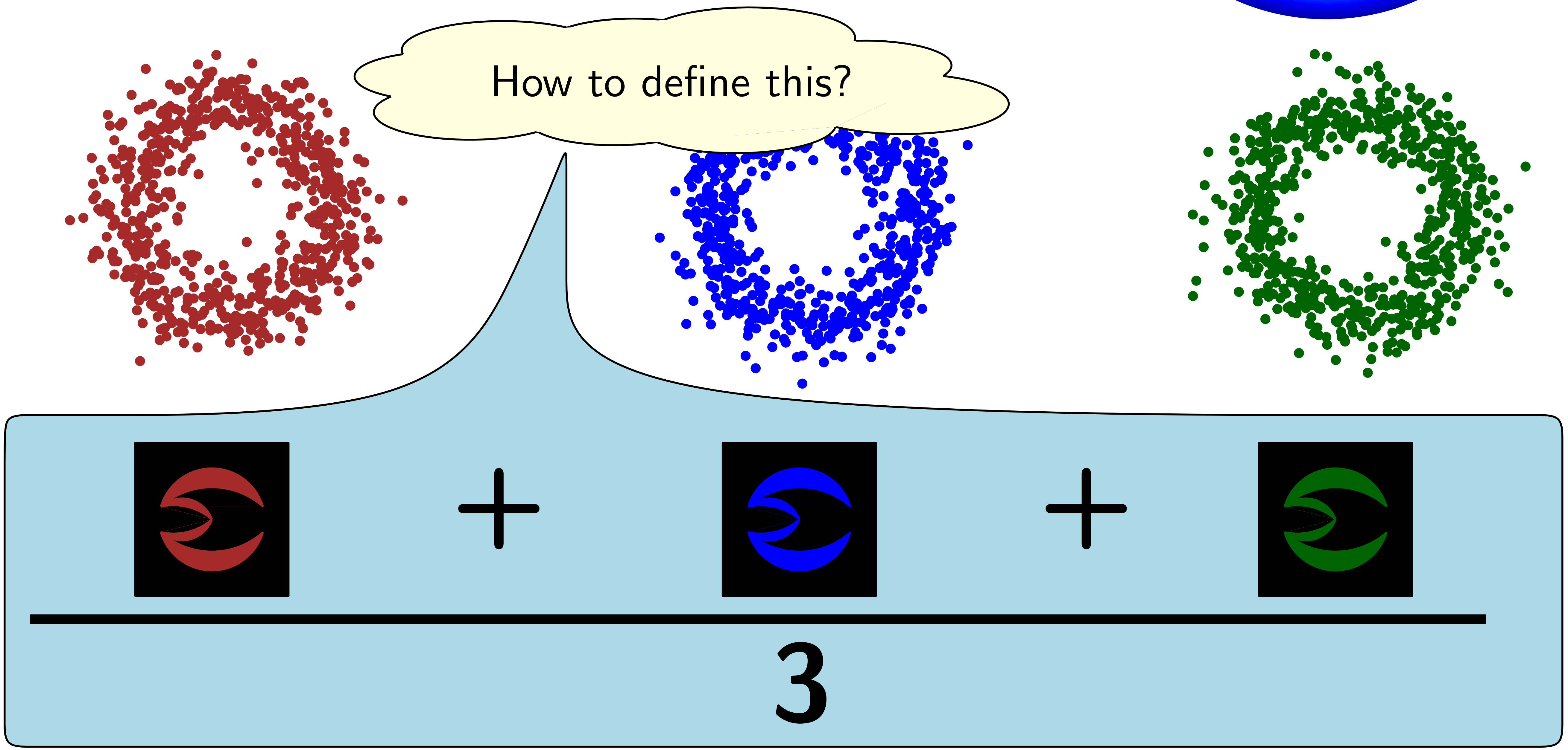


Specific Problem



Suppose we sample a circle with Gaussian noise



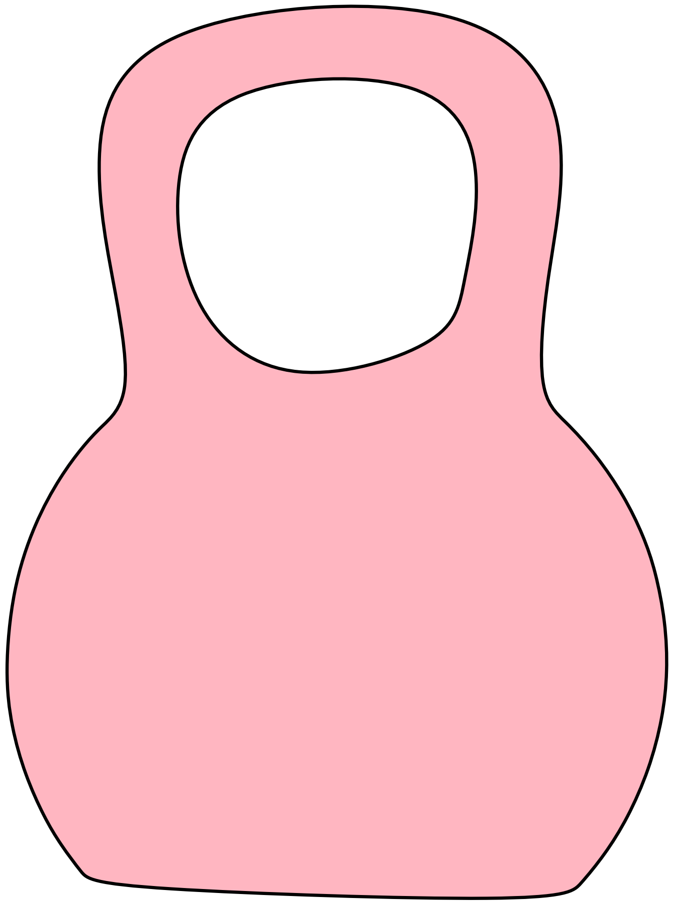


The distance map filtration

Typically, we represent data as a subset of some topological space, e.g. point cloud in $\mathbb{R}^n$, and associate to it a particular function, e.g. the distance to the point cloud.

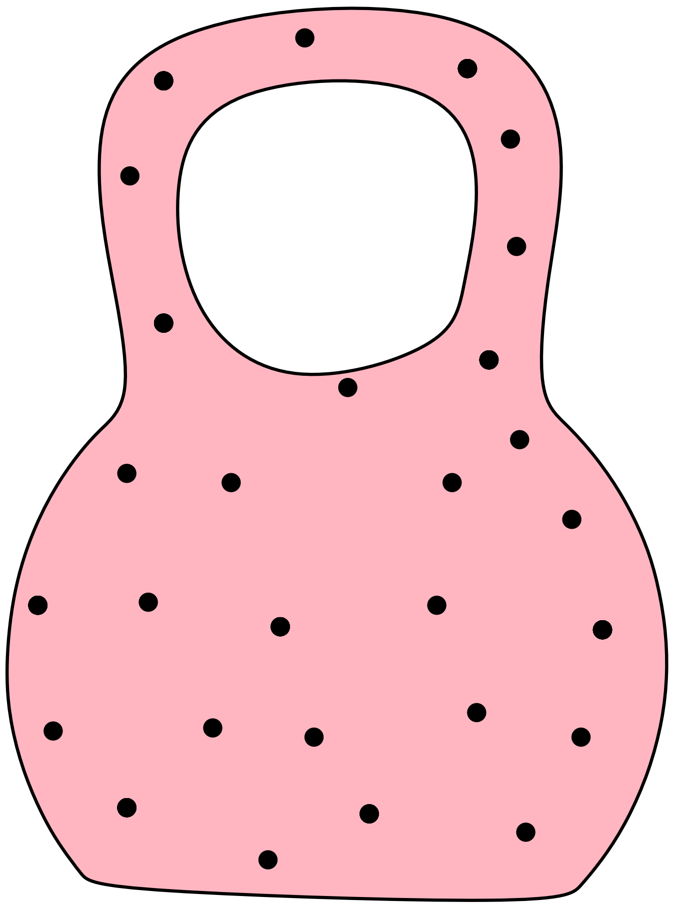
The distance map filtration

Typically, we represent data as a subset of some topological space, e.g. point cloud in $\mathbb{R}^n$, and associate to it a particular function, e.g. the distance to the point cloud.



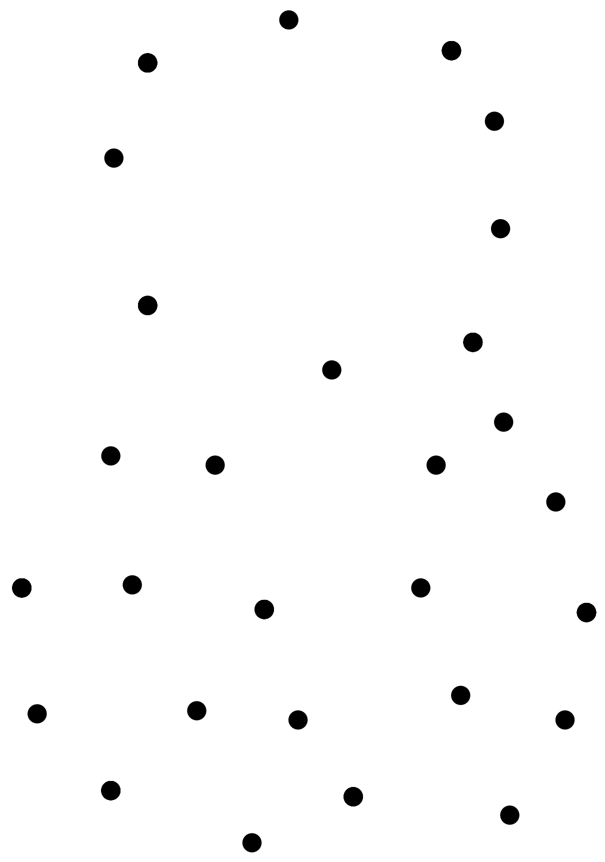
The distance map filtration

Typically, we represent data as a subset of some topological space, e.g. point cloud in $\mathbb{R}^n$, and associate to it a particular function, e.g. the distance to the point cloud.



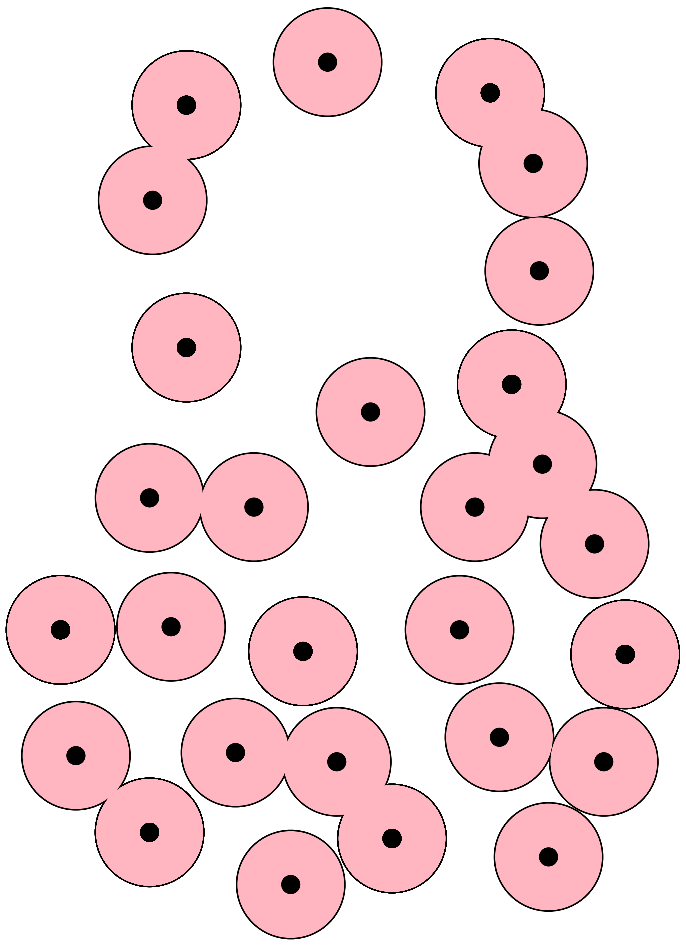
The distance map filtration

Typically, we represent data as a subset of some topological space, e.g. point cloud in $\mathbb{R}^n$, and associate to it a particular function, e.g. the distance to the point cloud.



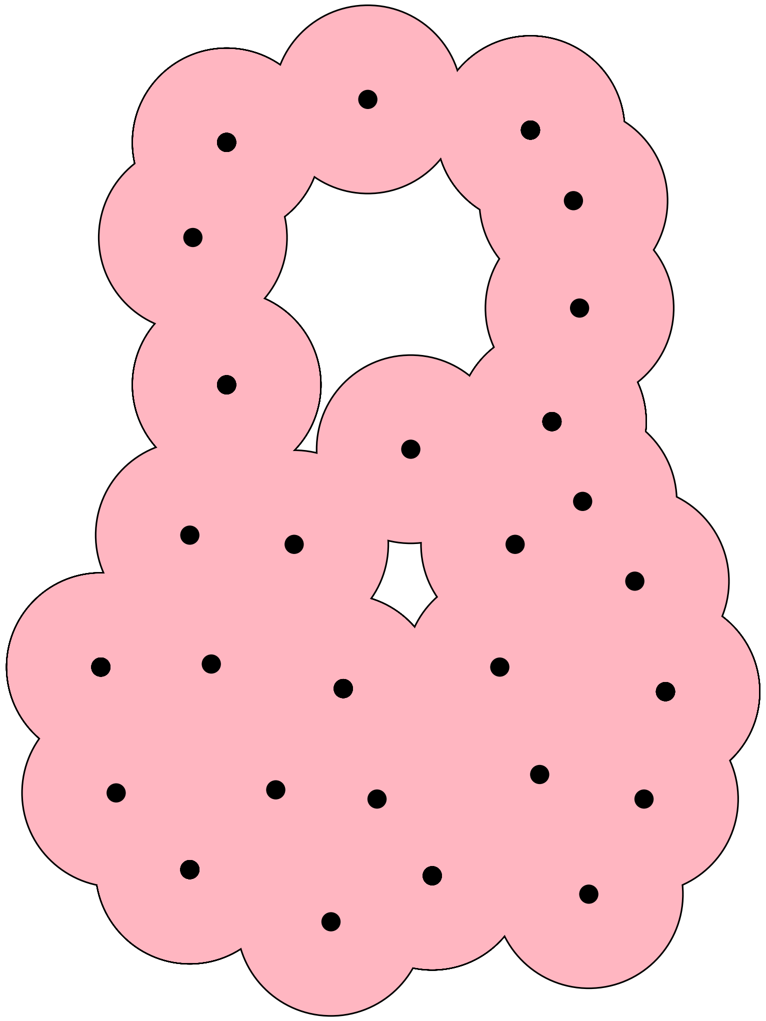
The distance map filtration

Typically, we represent data as a subset of some topological space, e.g. point cloud in $\mathbb{R}^n$, and associate to it a particular function, e.g. the distance to the point cloud.



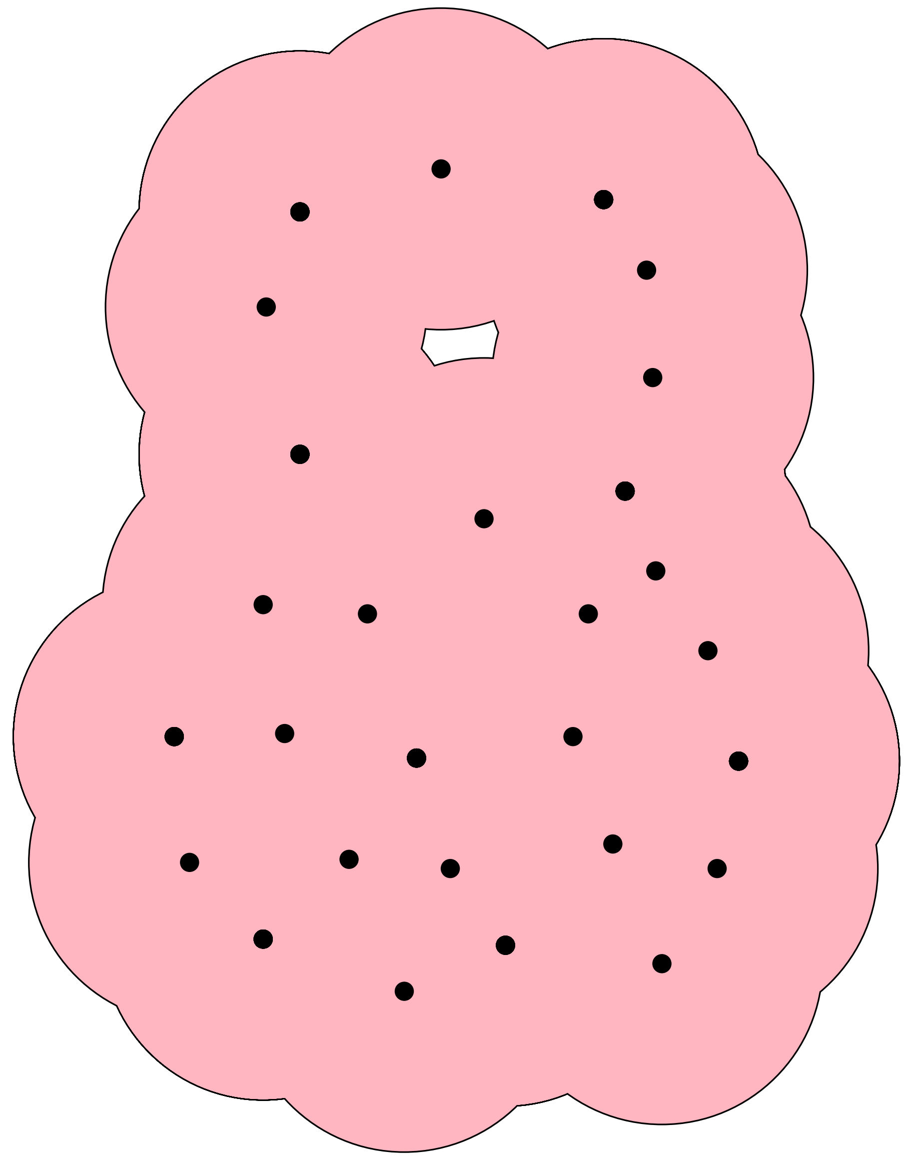
The distance map filtration

Typically, we represent data as a subset of some topological space, e.g. point cloud in $\mathbb{R}^n$, and associate to it a particular function, e.g. the distance to the point cloud.



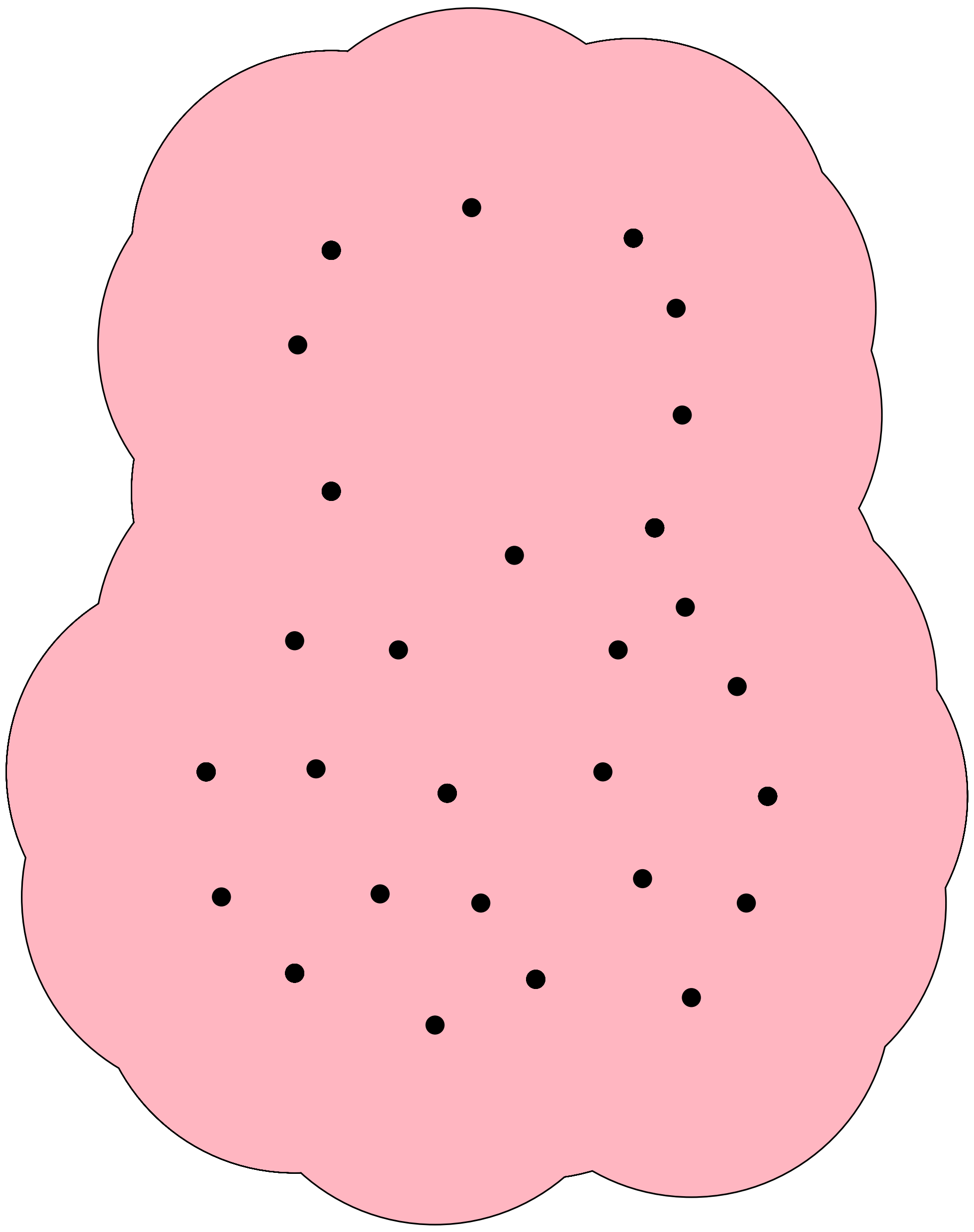
The distance map filtration

Typically, we represent data as a subset of some topological space, e.g. point cloud in $\mathbb{R}^n$, and associate to it a particular function, e.g. the distance to the point cloud.



The distance map filtration

Typically, we represent data as a subset of some topological space, e.g. point cloud in $\mathbb{R}^n$, and associate to it a particular function, e.g. the distance to the point cloud.



Persistence Idea

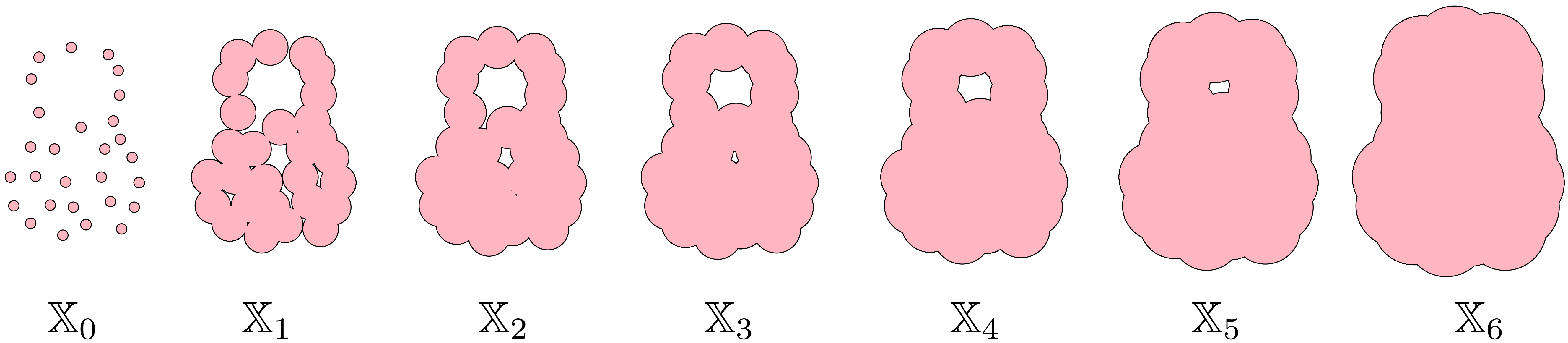
We have a nested family of topological spaces, i.e. a filtration:

$$\mathbb{X}_0 \subset \mathbb{X}_1 \subset \cdots \subset \mathbb{X}_n$$

Persistence Idea

We have a nested family of topological spaces, i.e. a filtration:

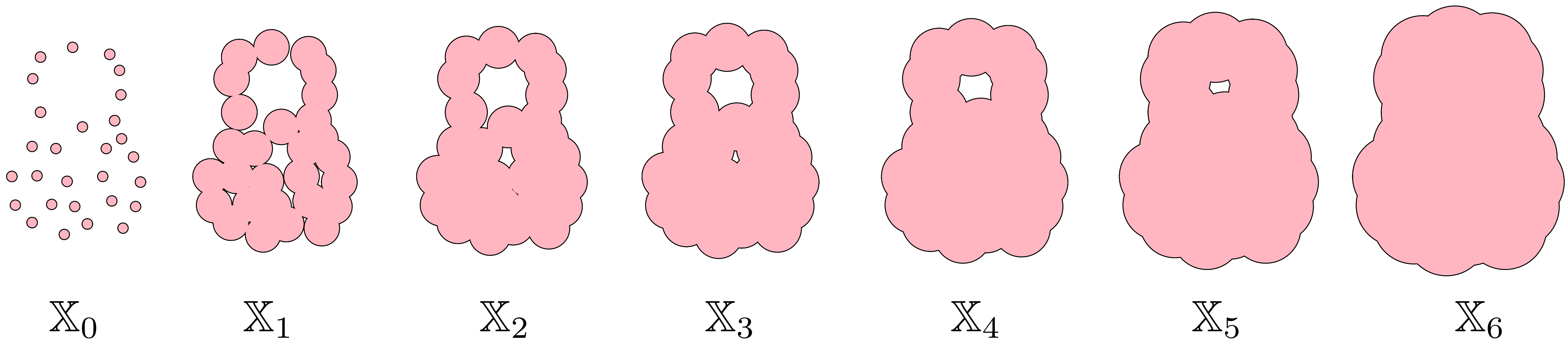
$$\mathbb{X}_0 \subset \mathbb{X}_1 \subset \cdots \subset \mathbb{X}_n$$



Persistence Idea

We have a nested family of topological spaces, i.e. a filtration:

$$\mathbb{X}_0 \subset \mathbb{X}_1 \subset \cdots \subset \mathbb{X}_n$$



$\mathbb{X}_0$ – bunch of connected components.

$\mathbb{X}_1$ – only one connected component is left, and seven 1-cycles are born.

$\mathbb{X}_2$ – five out of the seven 1-cycles died, but one new 1-cycle is born.

$\mathbb{X}_3$ – the new 1-cycle dies.

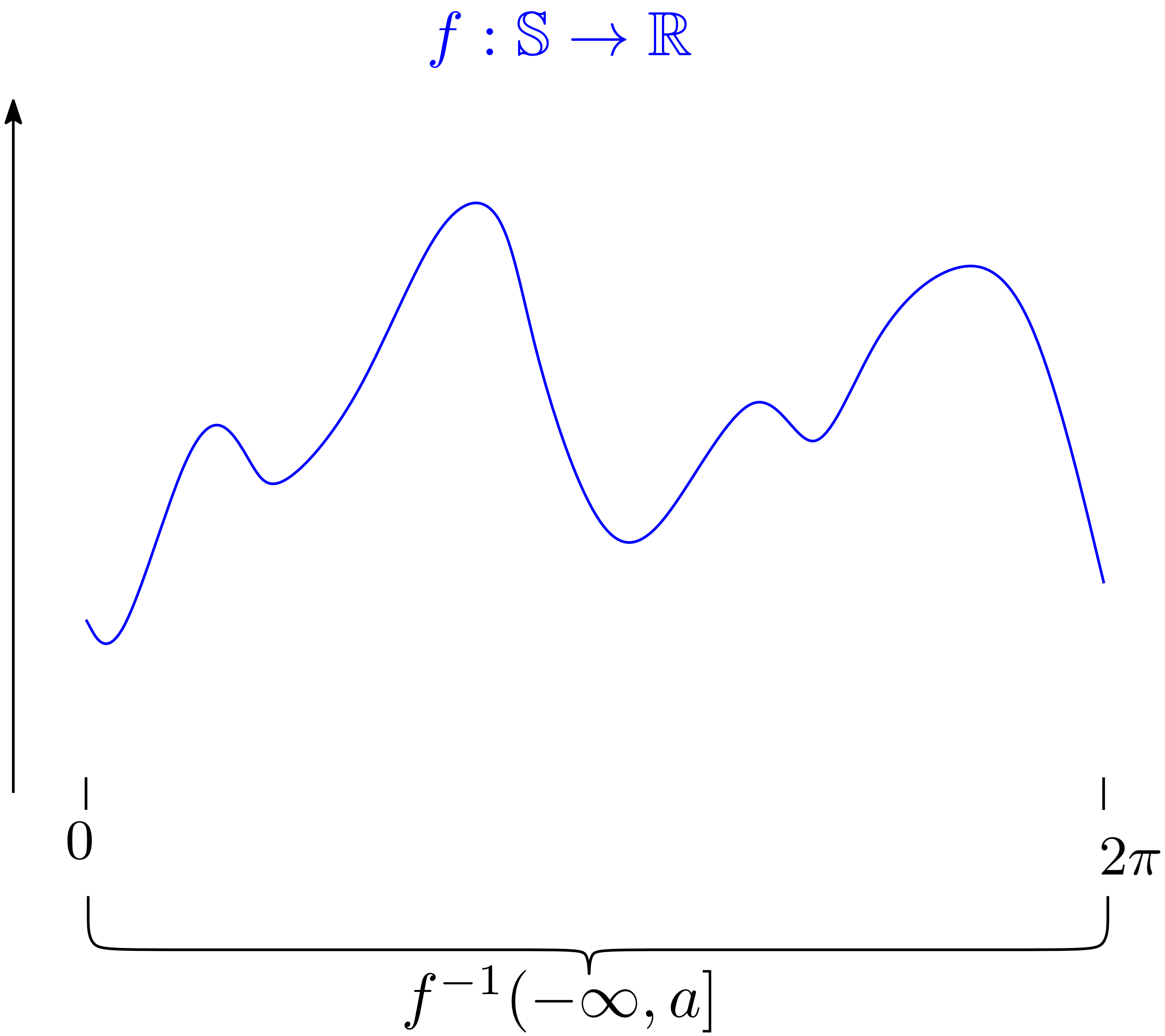
$\mathbb{X}_4$ – one of the oldest 1-cycles dies, one more keeps going.

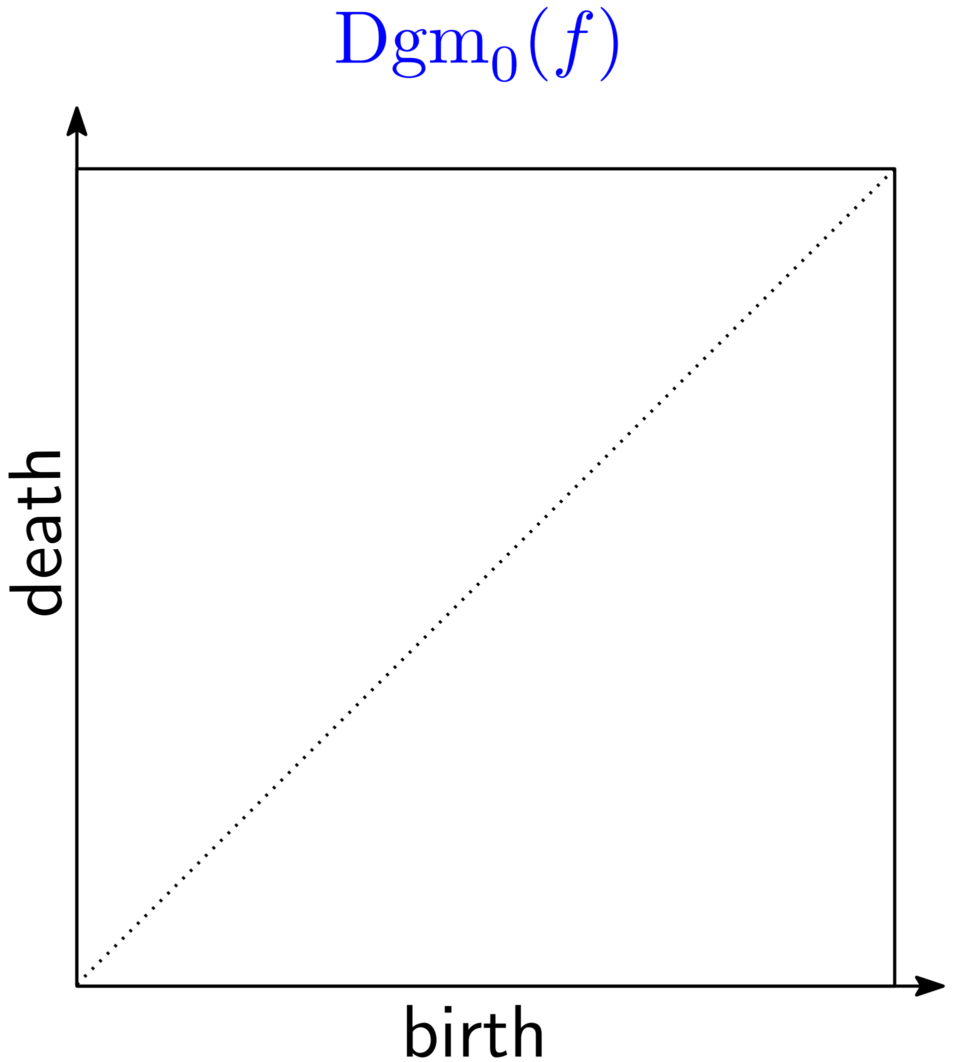
$\mathbb{X}_5$ – going strong!

$\mathbb{X}_6$ – everybody's dead!

Persistence Diagrams

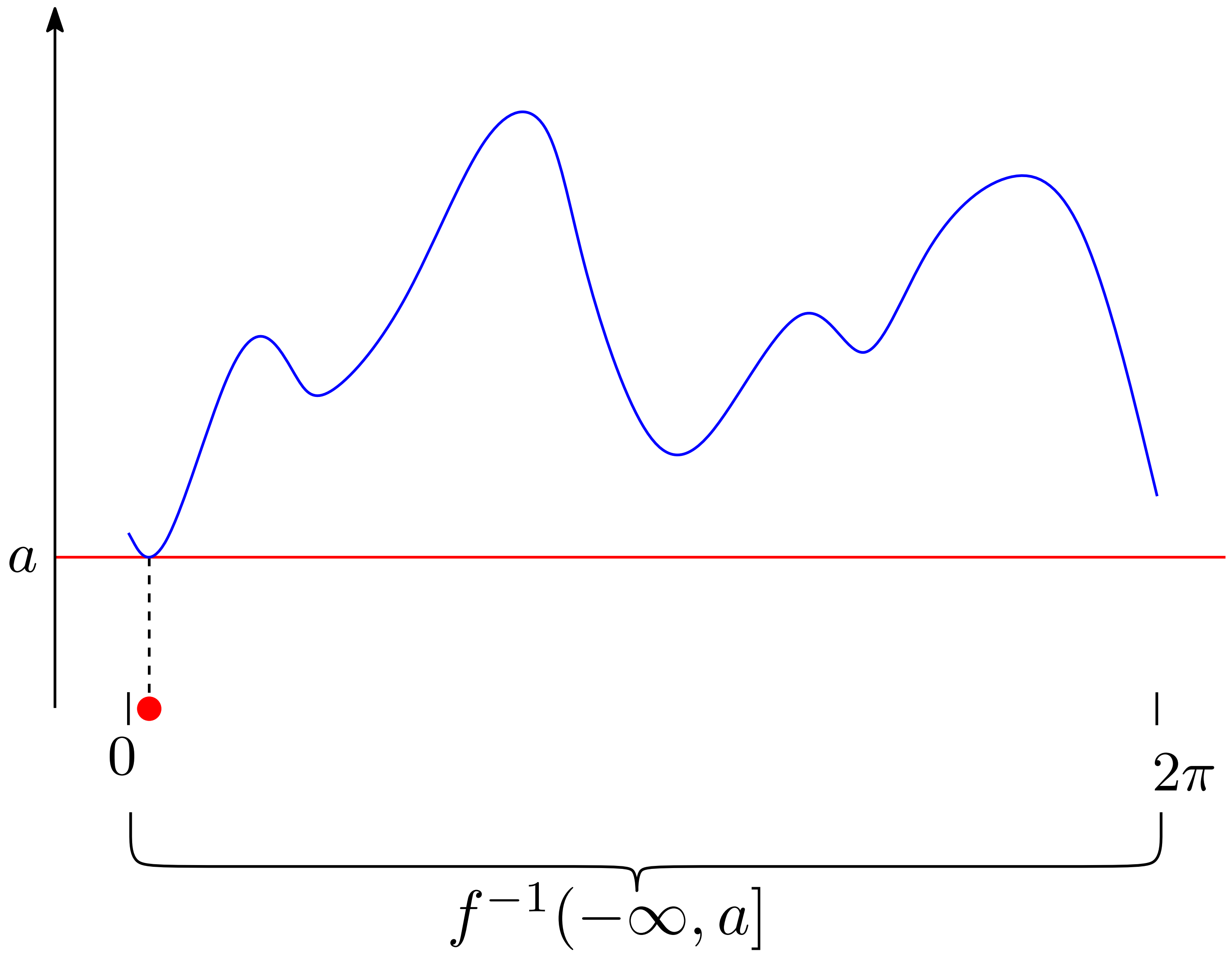
Persistence Diagrams



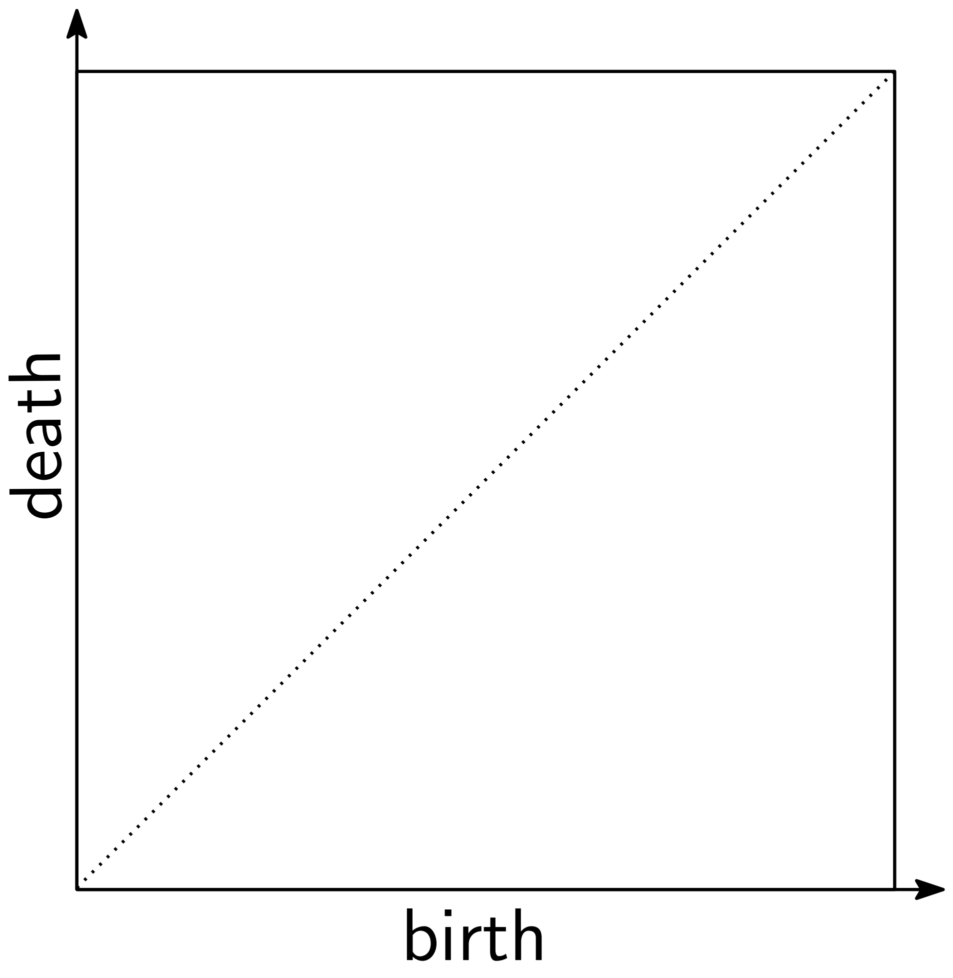


Persistence Diagrams

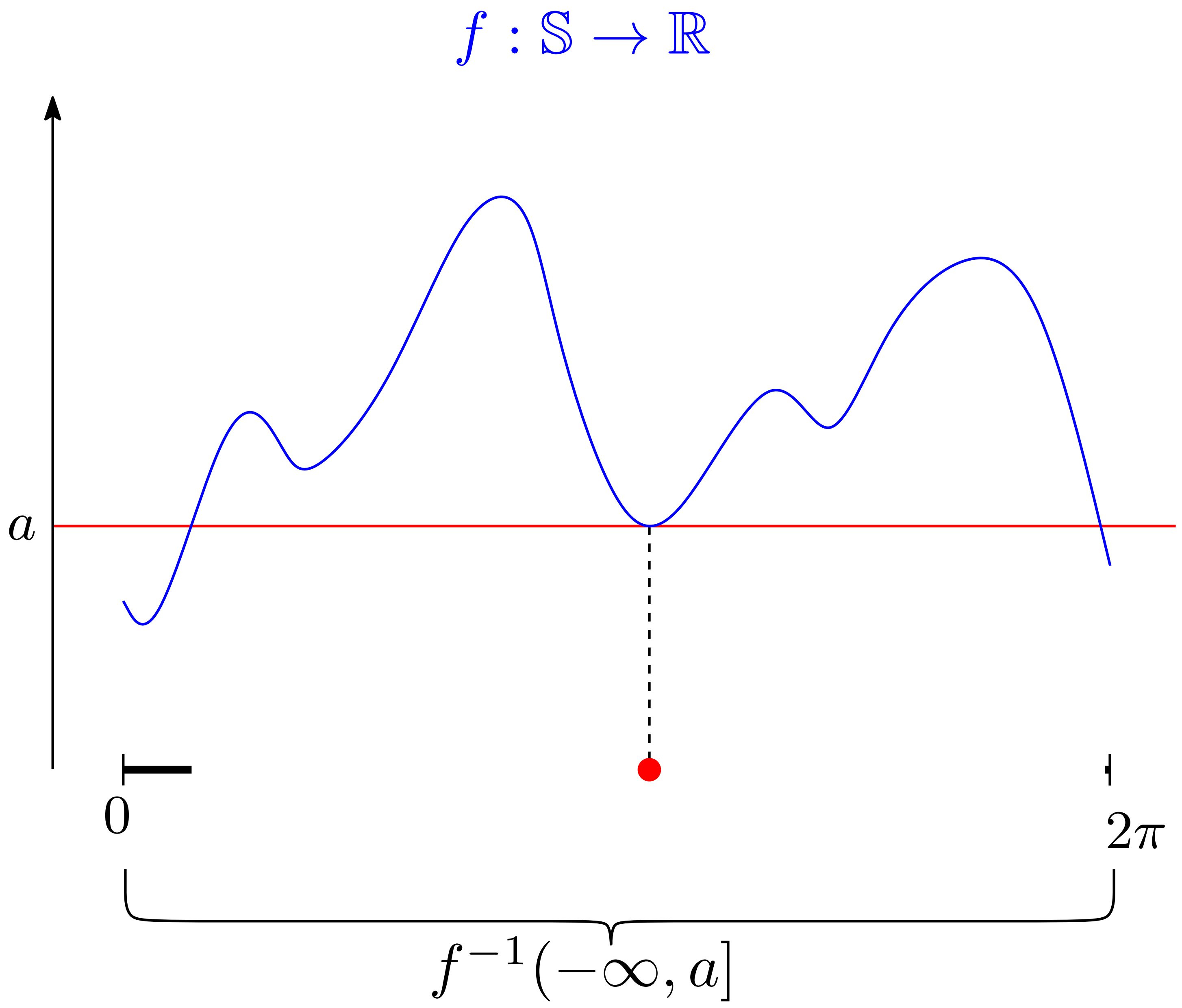
$f : \mathbb{S} \to \mathbb{R}$

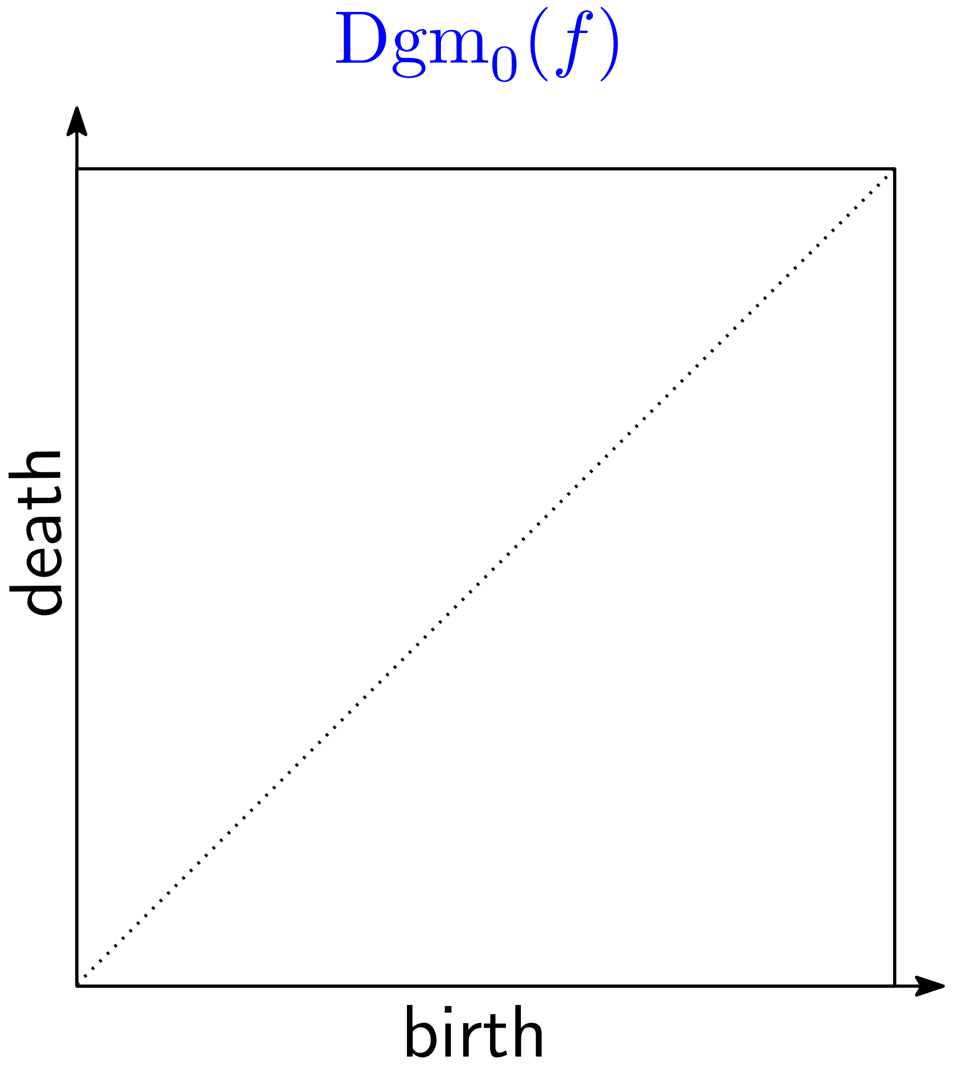


$\mathrm{Dgm}_0(f)$

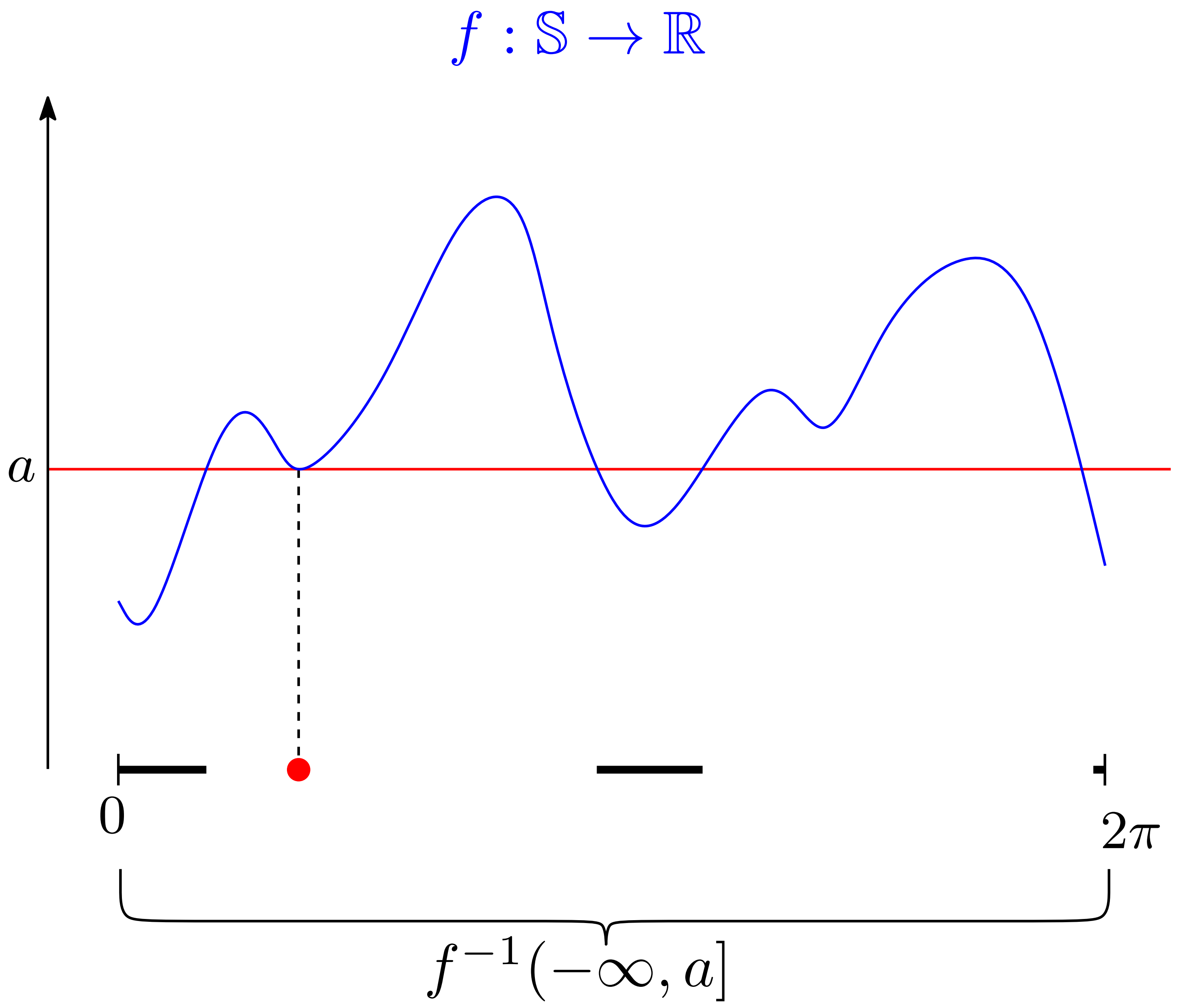


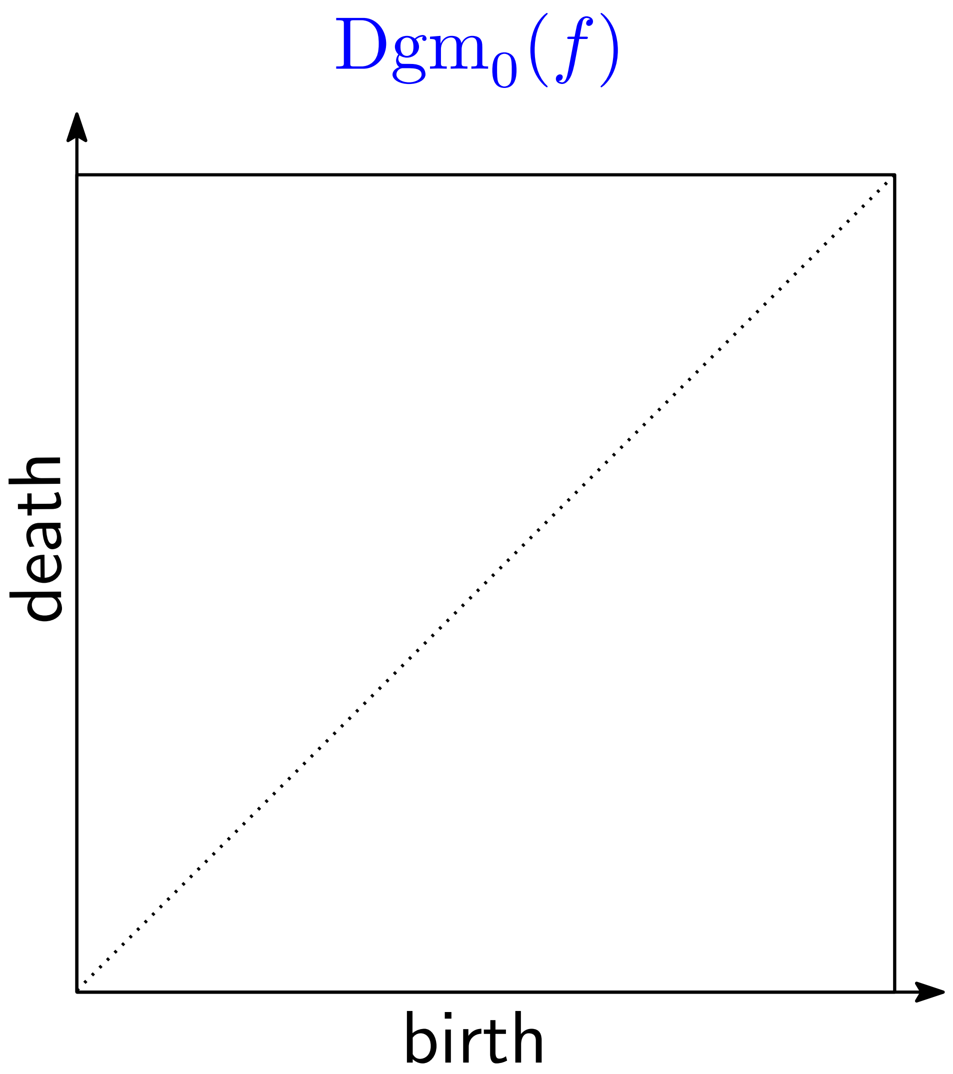
Persistence Diagrams



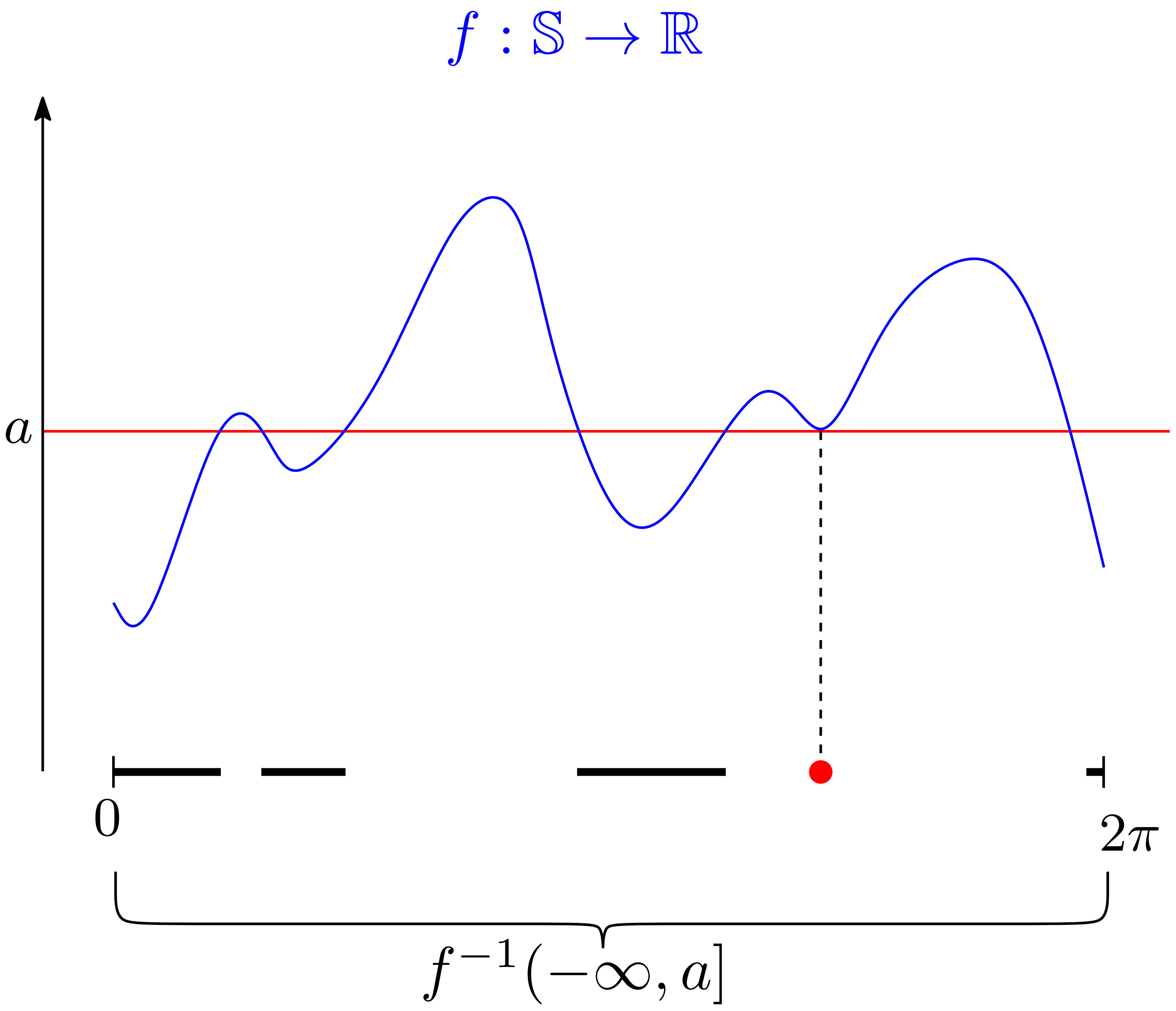


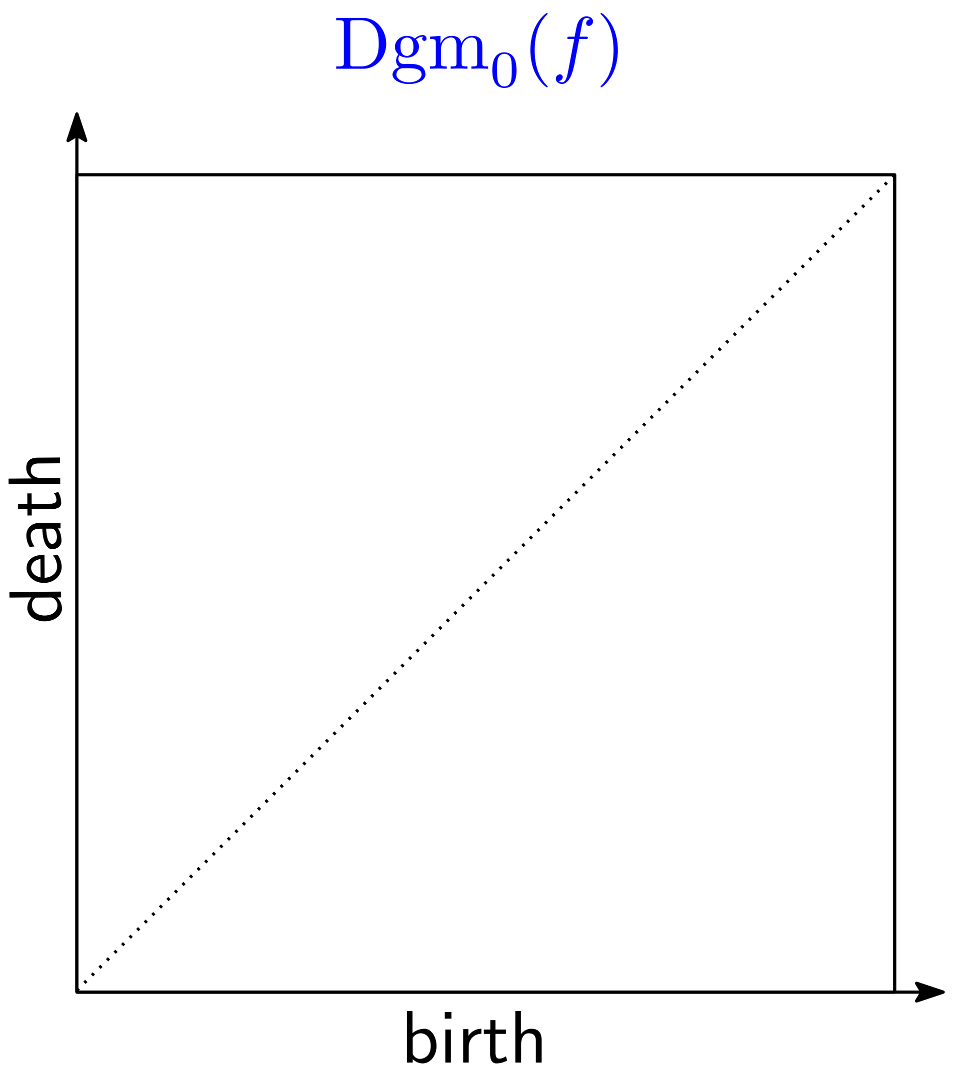
Persistence Diagrams



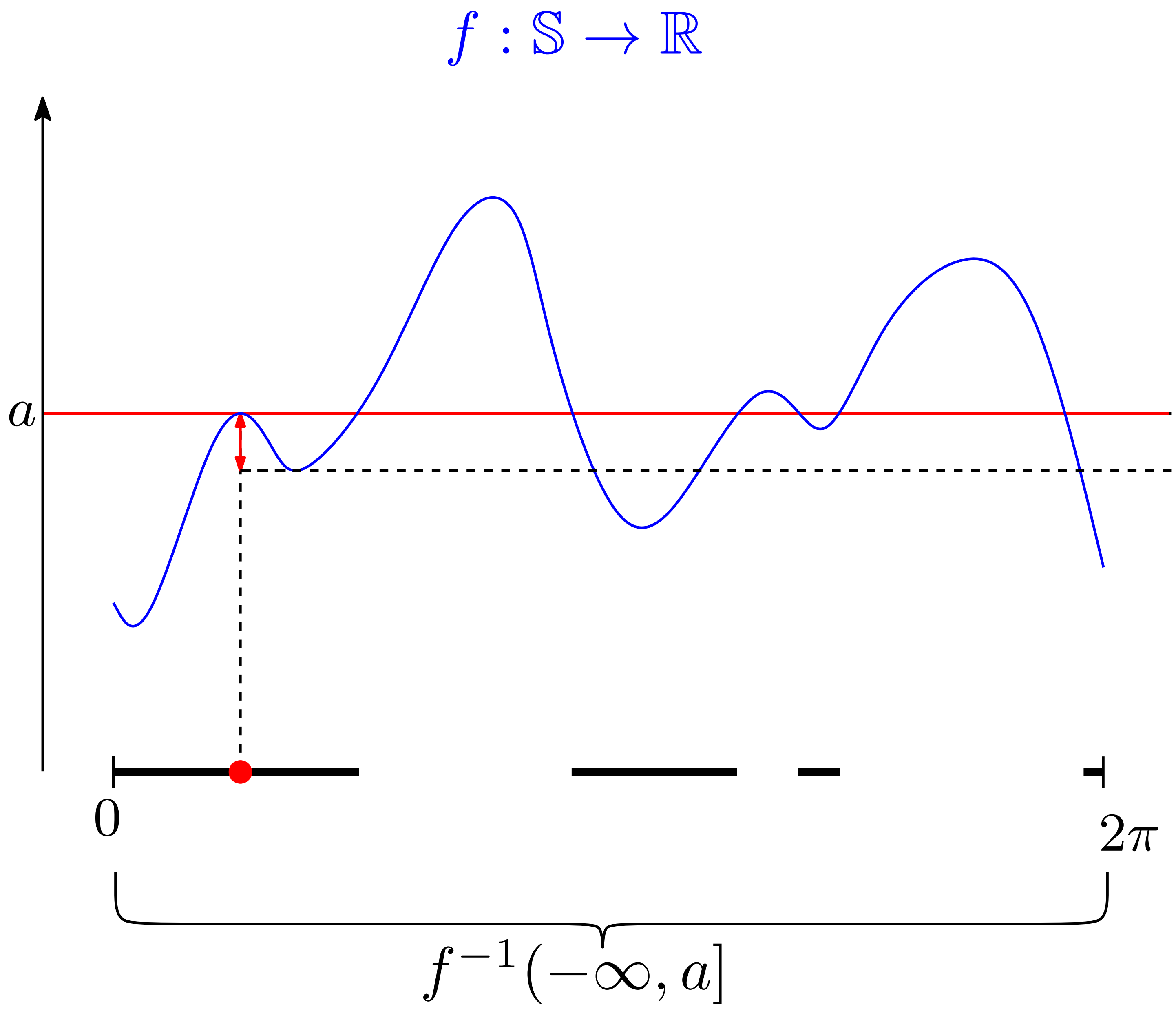


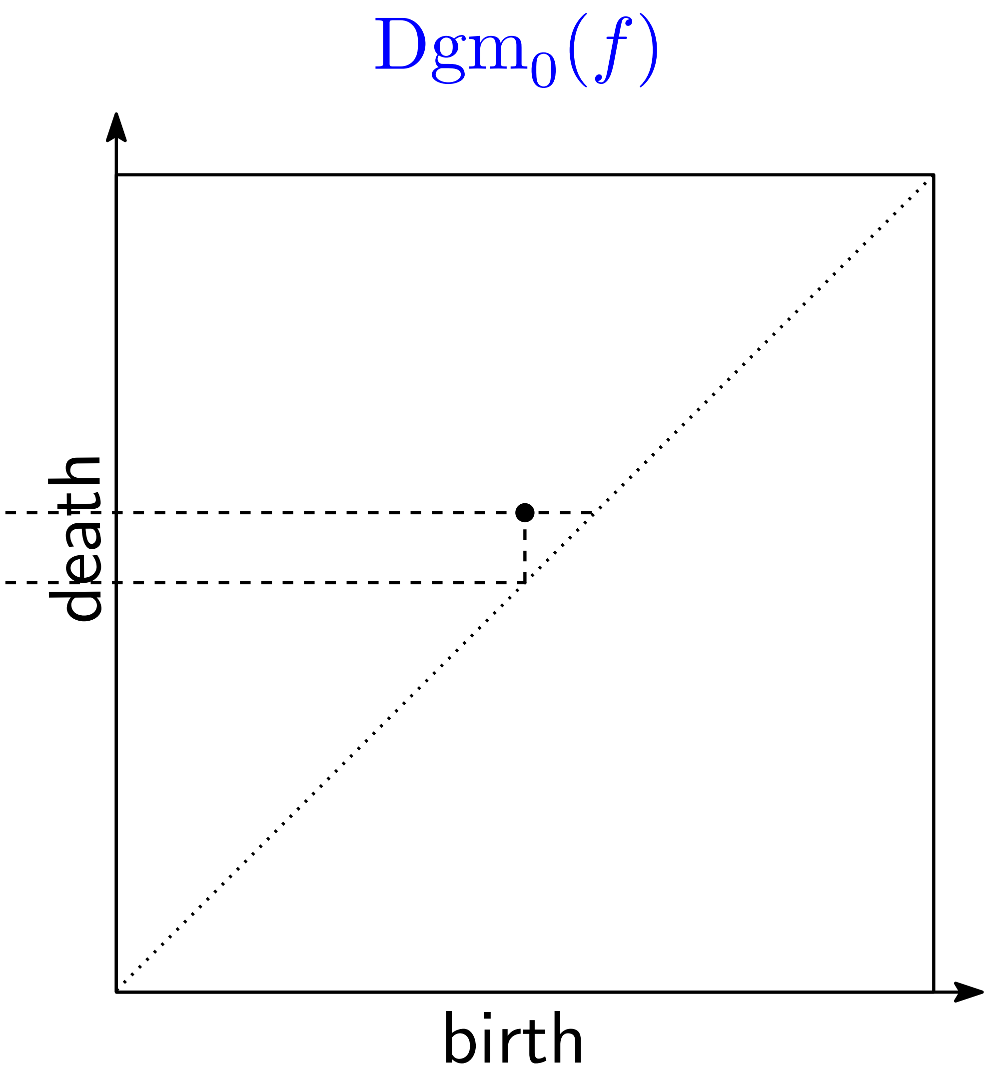
Persistence Diagrams



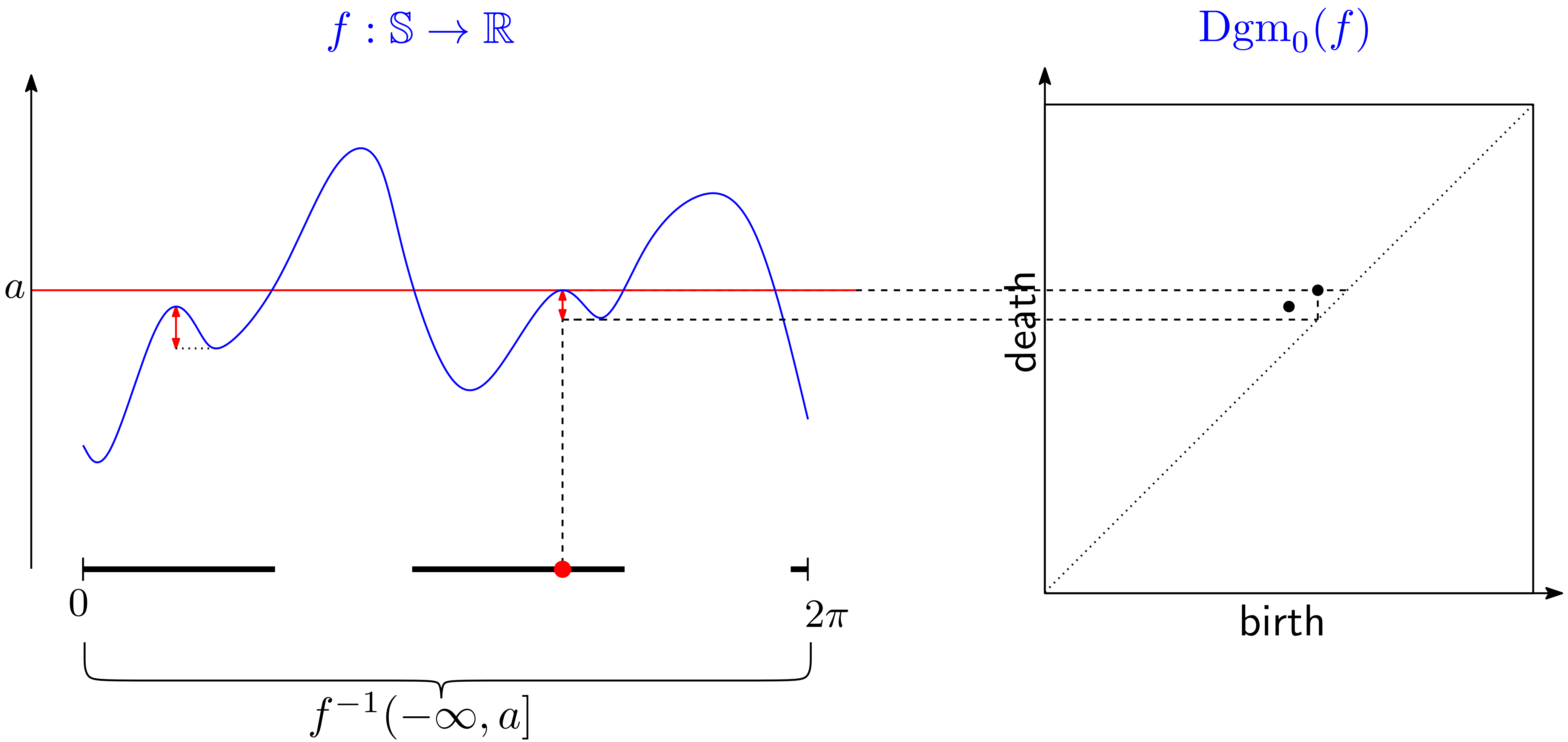


Persistence Diagrams

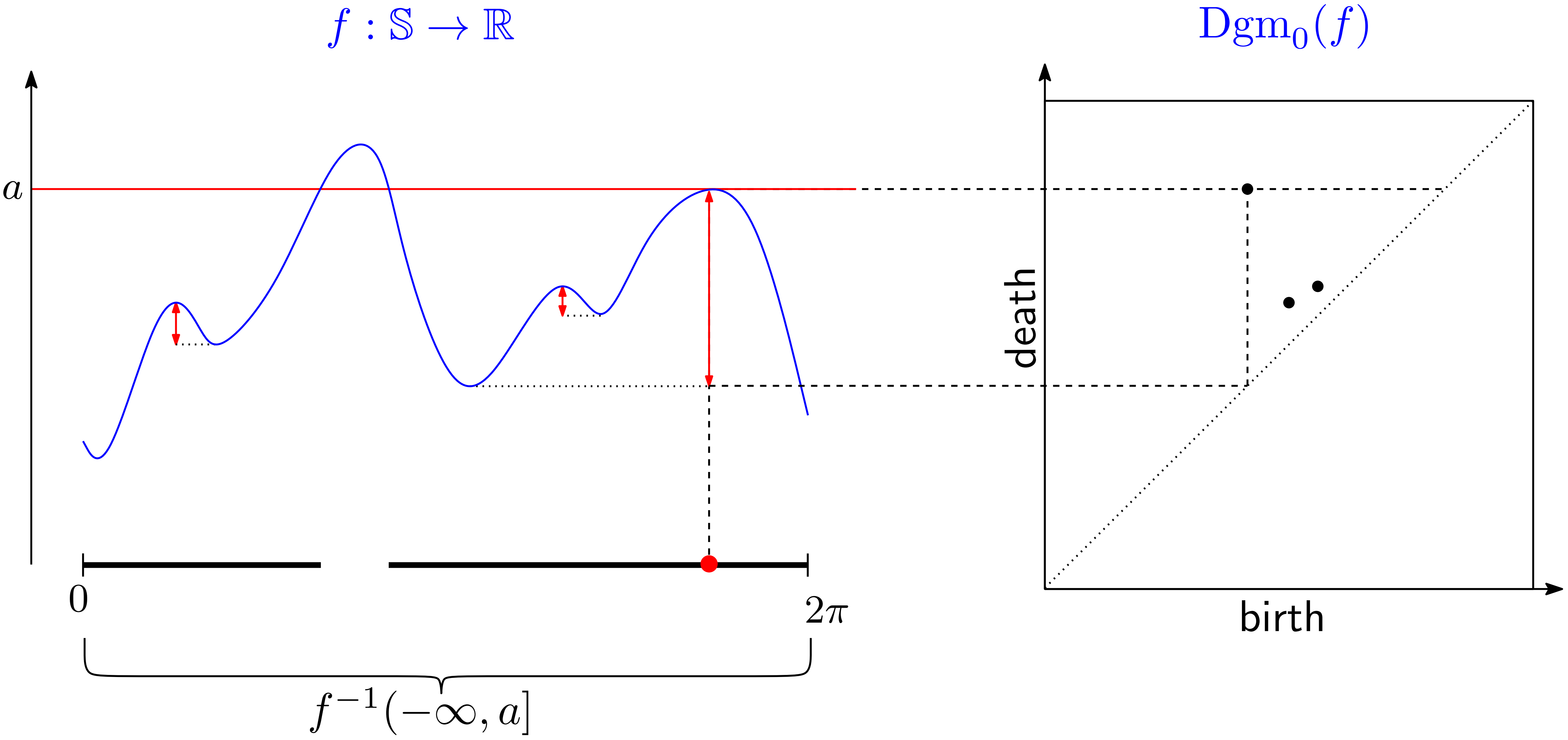




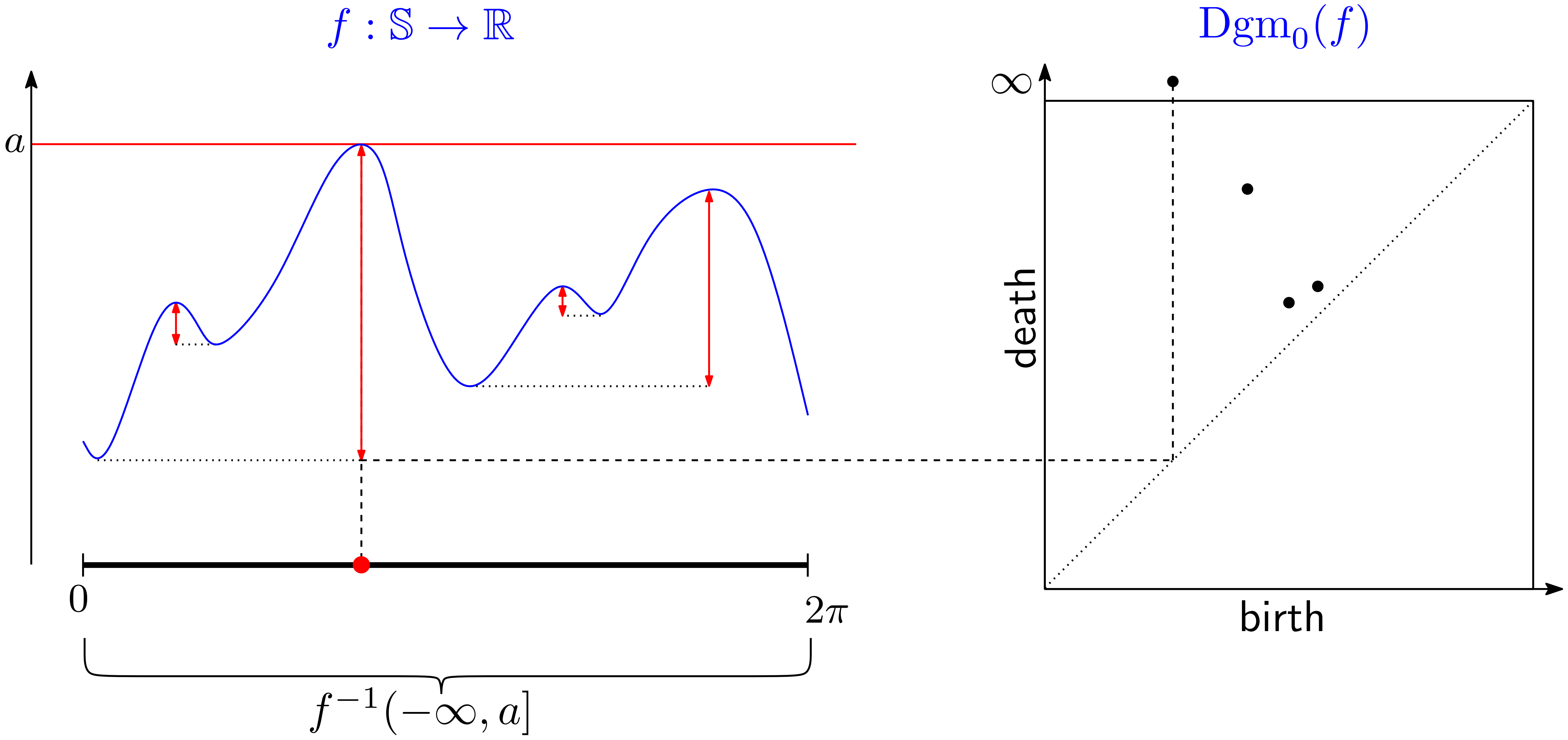
Persistence Diagrams



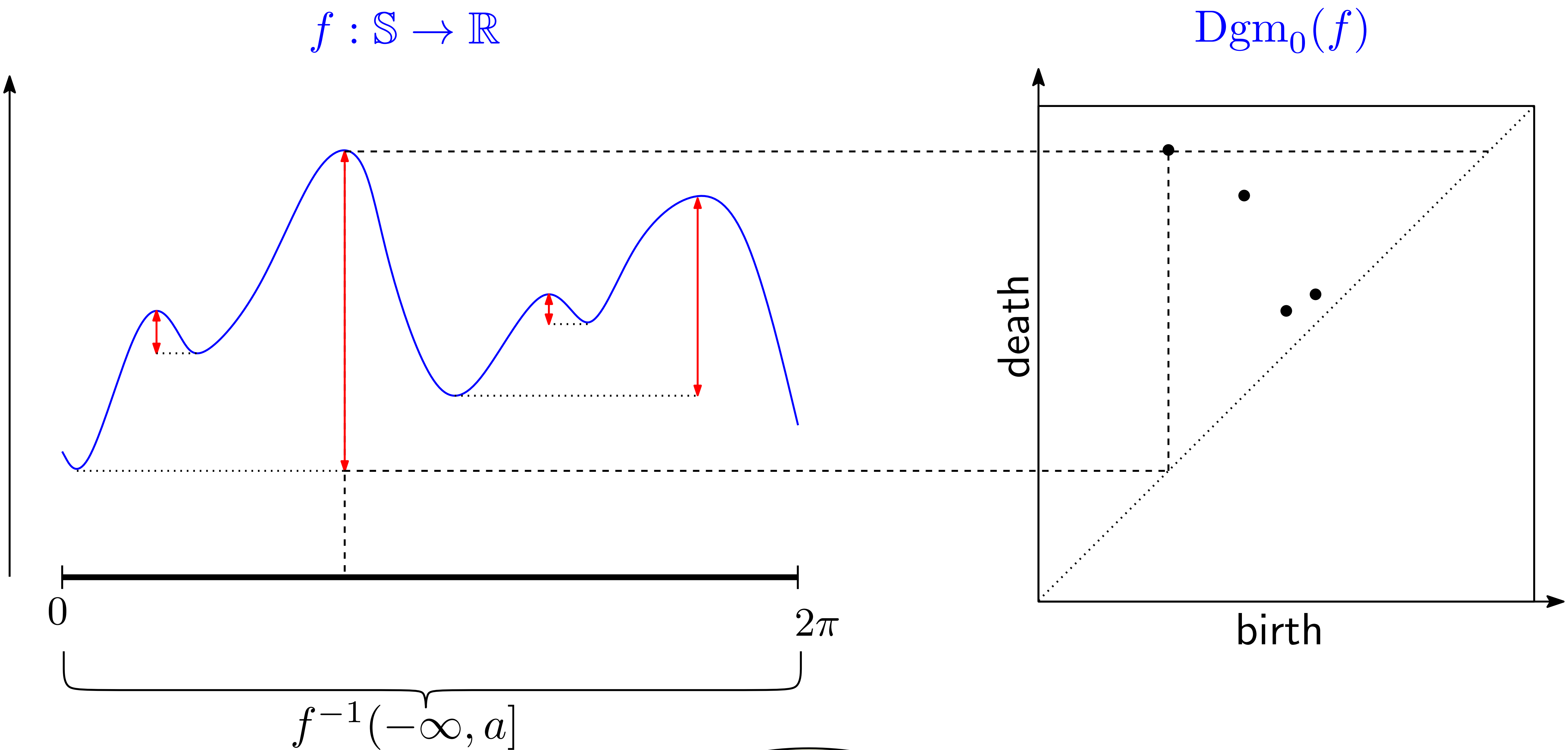
Persistence Diagrams



Persistence Diagrams

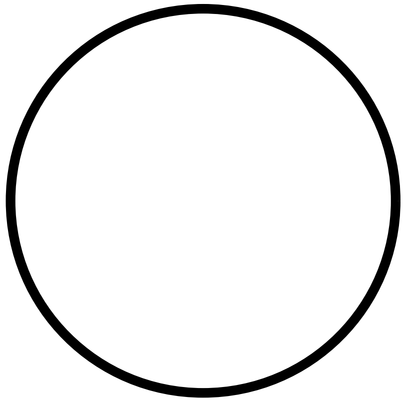


Persistence Diagrams

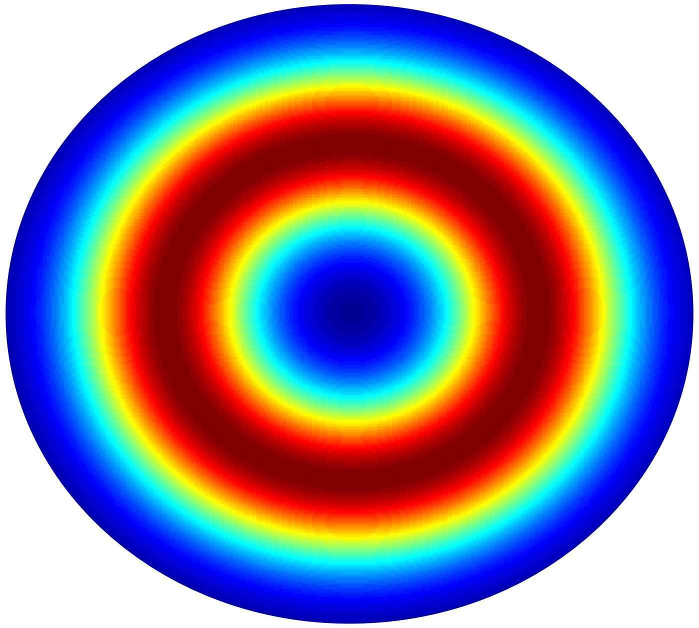


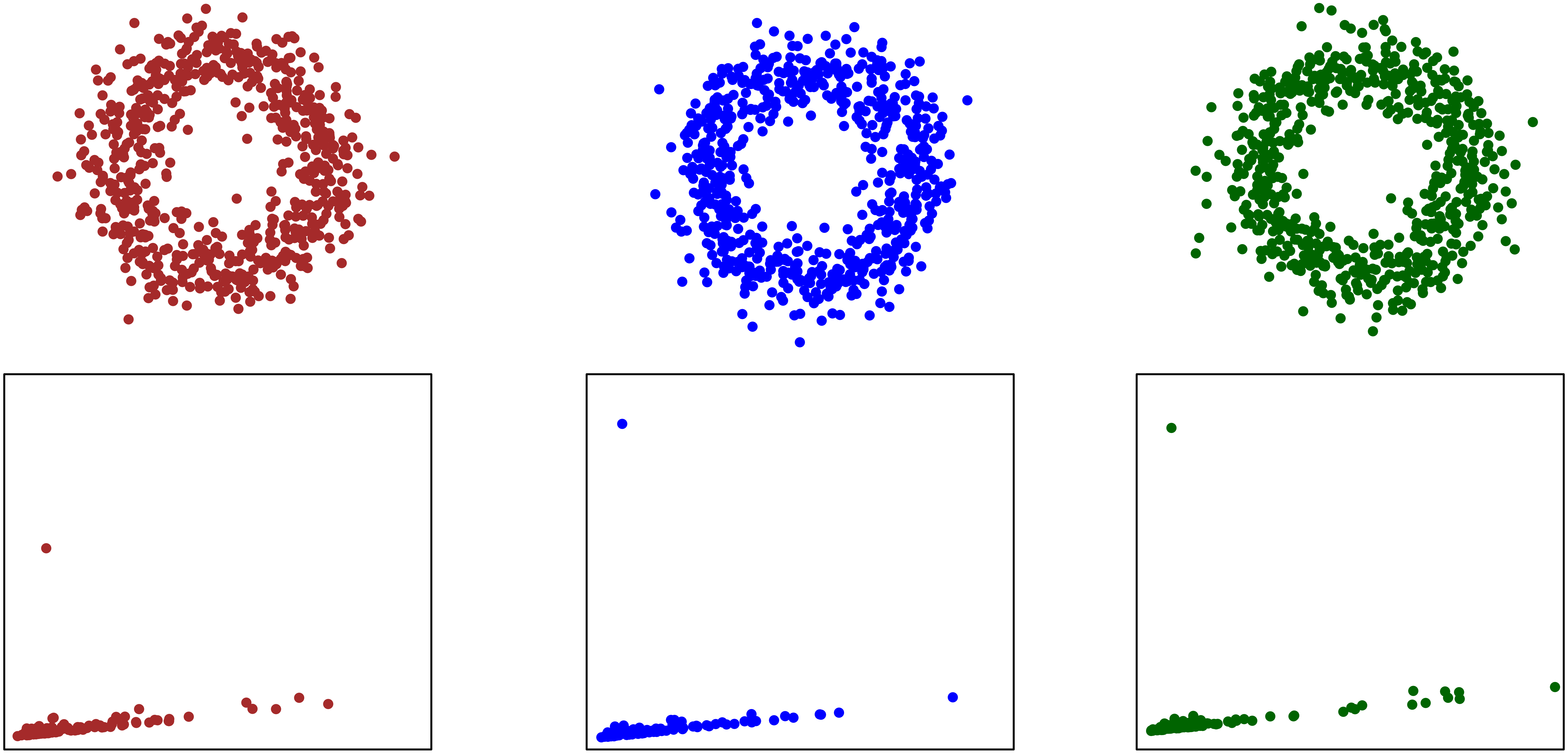
For a compact space $\mathbb{X}$ it is convenient to kill immortal classes at $\max_{x \in \mathbb{X}} f(x)$

Very Specific Problem

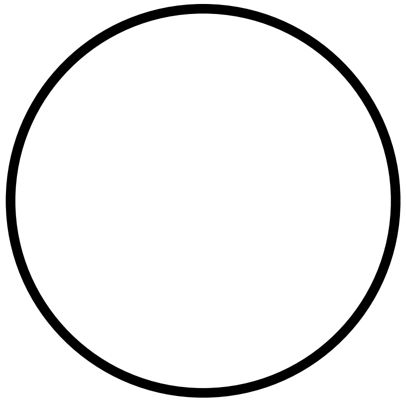


Suppose we sample a circle with Gaussian noise

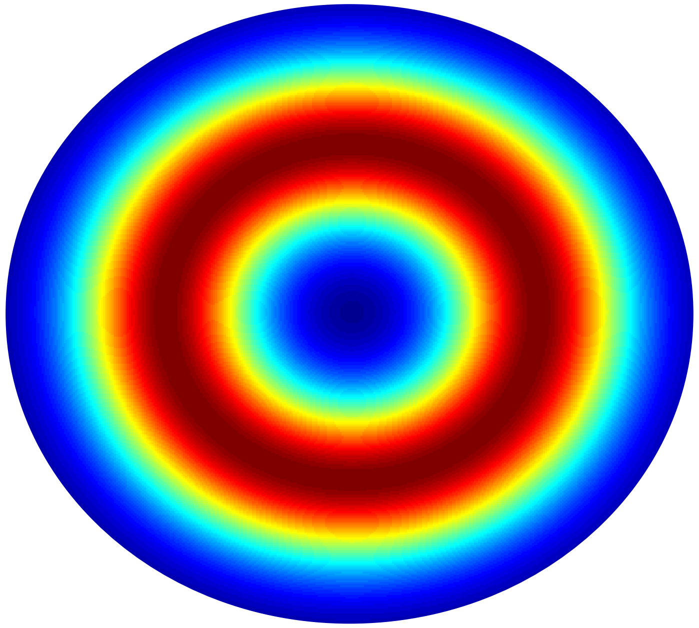




Very Specific Problem

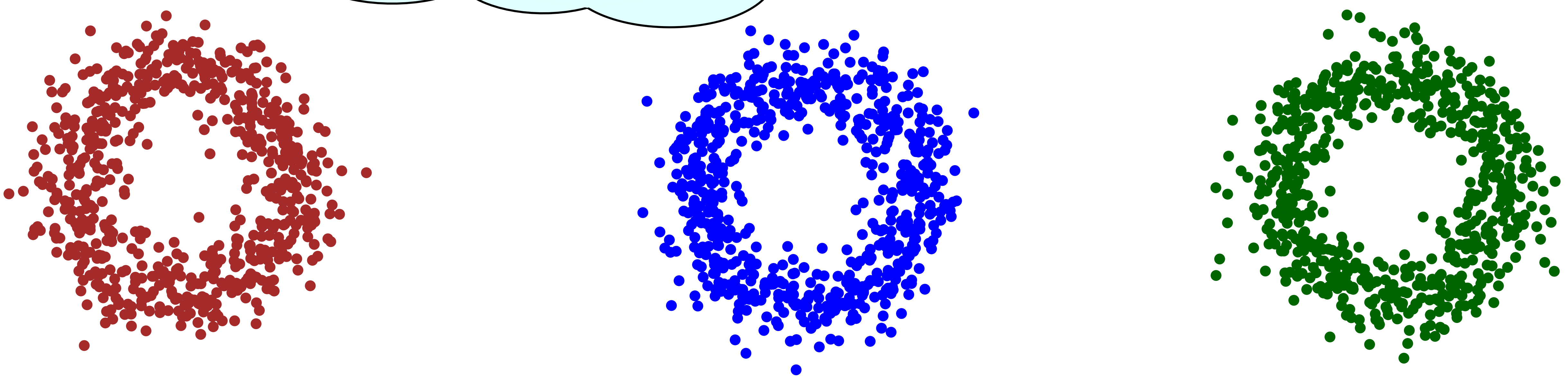


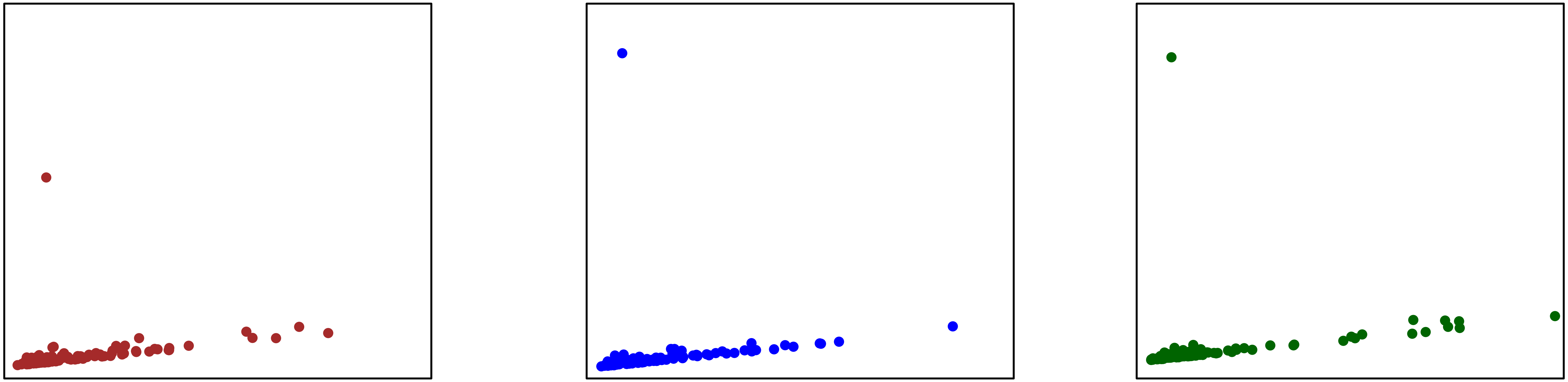
Suppose we sample a circle with Gaussian noise



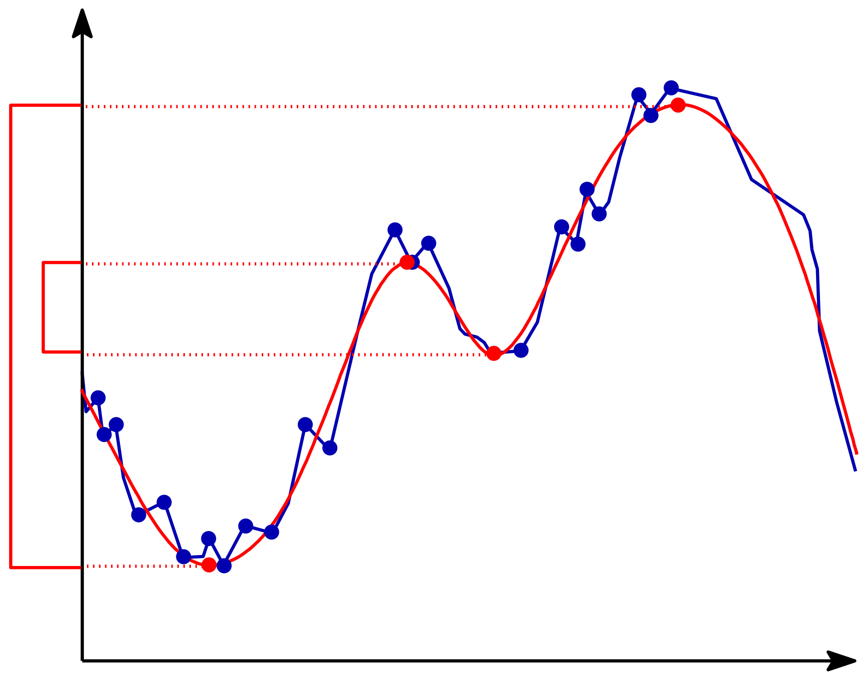
What's the mean?

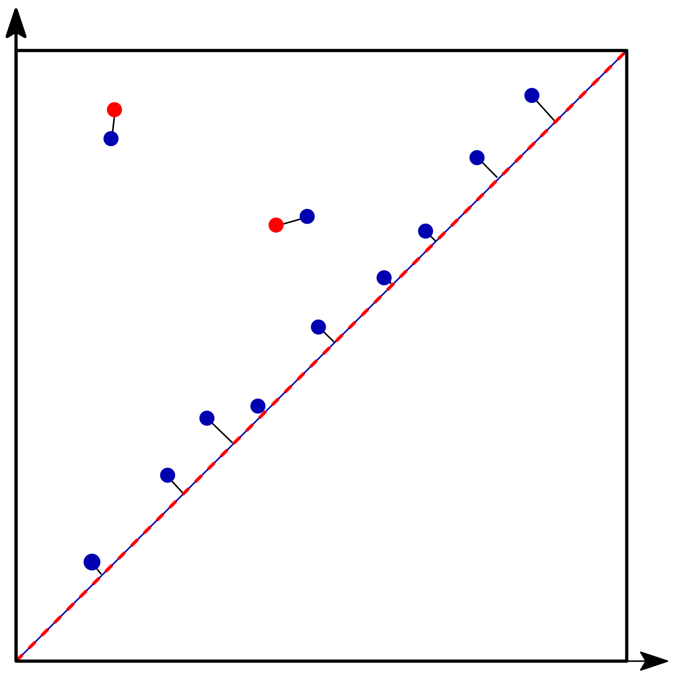
What's the variance?





Metric on Persistence Diagrams





Metric on Persistence Diagrams

Wasserstein Distance

The p-th Wasserstein distance between the persistence diagrams of f and g is

$$W_p(f,g) = \left[\sum_{x \in \mathrm{Dgm}_\ell(f)} \|x - \gamma_\ell(x)\|_\infty^p \right]^{\frac{1}{p}},$$

where the infimum is over all bijections $\gamma_\ell : \mathrm{Dgm}_\ell(f) \to \mathrm{Dgm}_\ell(g)$.

Metric on Persistence Diagrams

WASSERSTEIN DISTANCE

The p-th Wasserstein distance between the persistence diagrams of f and g is

$$W_p(f,g) = \left[\sum_{x \in \mathrm{Dgm}_\ell(f)} \|x - \gamma_\ell(x)\|_\infty^p \right]^{\frac{1}{p}},$$

where the infimum is over all bijections $\gamma_\ell : \mathrm{Dgm}_\ell(f) \to \mathrm{Dgm}_\ell(g)$.

WASSERSTEIN STABILITY THEOREM[*Cohen-Steiner, Edelsbrunner, Harer, M.*].
Under mild conditions on a metric space $\mathbb{X}$ and functions $f, g : \mathbb{X} \to \mathbb{R}$ we have

$$W_p(f,g) \leq C \|f - g\|_\infty^{1 - \frac{k}{p}}$$

Space of Persistence Diagrams

The set of persistence diagrams (with a finite number of points in them) endowed with the Wassertstein distance is a metric space.

Space of Persistence Diagrams

The set of persistence diagrams (with a finite number of points in them) endowed with the Wassertstein distance is a metric space.

But it is not complete!

Consider $x_n = (0, 2^{-n}) \in \mathbb{R}^2$, $n \in \mathbb{N}$, and let d_n be the persistence diagram containing $x_1, \ldots, x_n$ (each with multiplicity 1). Then

$$W_p(d_n, d_{n+k}) < \frac{1}{2^n},$$

so d_n is Cauchy. It is clear, however, that the number of off-diagonal points in d_n grows to ∞ as $n \to \infty$, so this sequence cannot have a limit in our space.

Space of Persistence Diagrams

The set of persistence diagrams (with a finite number of points in them) endowed with the Wassertstein distance is a metric space.

But it is not complete!

Consider $x_n = (0, 2^{-n}) \in \mathbb{R}^2$, $n \in \mathbb{N}$, and let d_n be the persistence diagram containing $x_1, \ldots, x_n$ (each with multiplicity 1). Then

$$W_p(d_n, d_{n+k}) < \frac{1}{2^n},$$

so d_n is Cauchy. It is clear, however, that the number of off-diagonal points in d_n grows to ∞ as $n \to \infty$, so this sequence cannot have a limit in our space.

Definition. A persistence diagram is a *countable* multiset of points in $\mathbb{R}^2$ along with the diagonal $\Delta = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$, where each point on the diagonal has infinite multiplicity.

Definition. The space of persistence diagrams is defined as

$$D_p = \{d | \mathrm{Pers}_p(d) < \infty\}, \quad \text{where} \quad \mathrm{Pers}_p(d) = \sum_{x \in d} \mathrm{pers}(x)^p$$

Space of Persistence Diagrams

THEOREM[*M., Mukherjee, Harer*] The space D_p is complete and separable.

Space of Persistence Diagrams

THEOREM[*M., Mukherjee, Harer*] The space D_p is complete and separable.

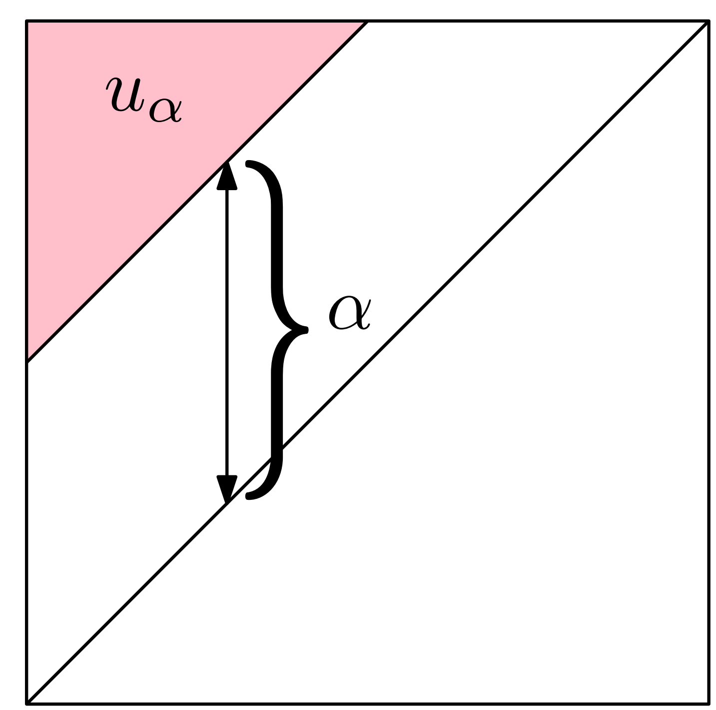
The proof of completeness of ℓ_p does not translate well here because the notion of a pointwise convergence is not defined.

Space of Persistence Diagrams

THEOREM[*M., Mukherjee, Harer*] The space D_p is complete and separable.

The proof of completeness of ℓ_p does not translate well here because the notion of a pointwise convergence is not defined.

Instead, we define *persistence-wise* convergence.



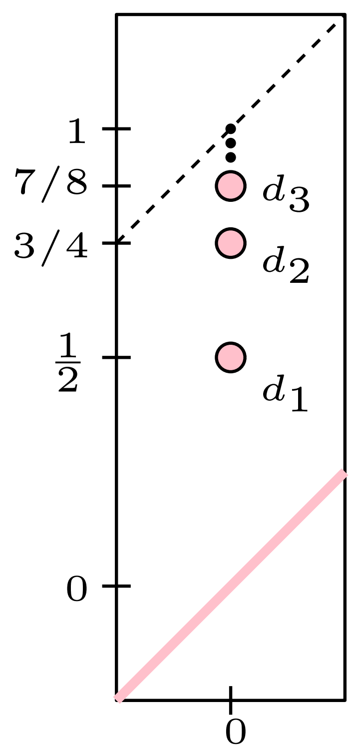
Space of Persistence Diagrams

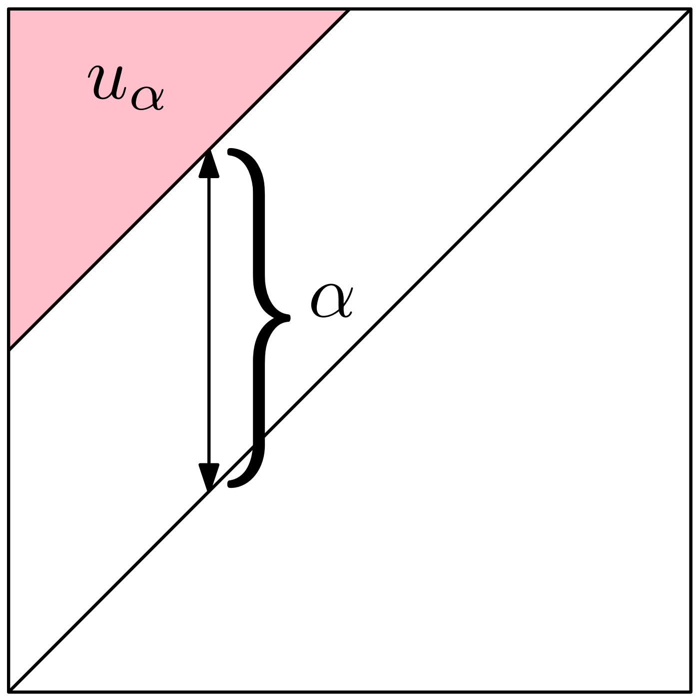
THEOREM[*M., Mukherjee, Harer*] The space D_p is complete and separable.

The proof of completeness of ℓ_p does not translate well here because the notion of a pointwise convergence is not defined.

Instead, we define *persistence-wise* convergence.

The problem is convergence from below.





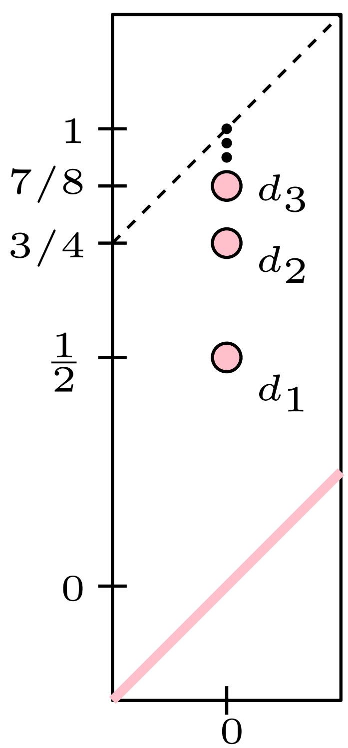
Space of Persistence Diagrams

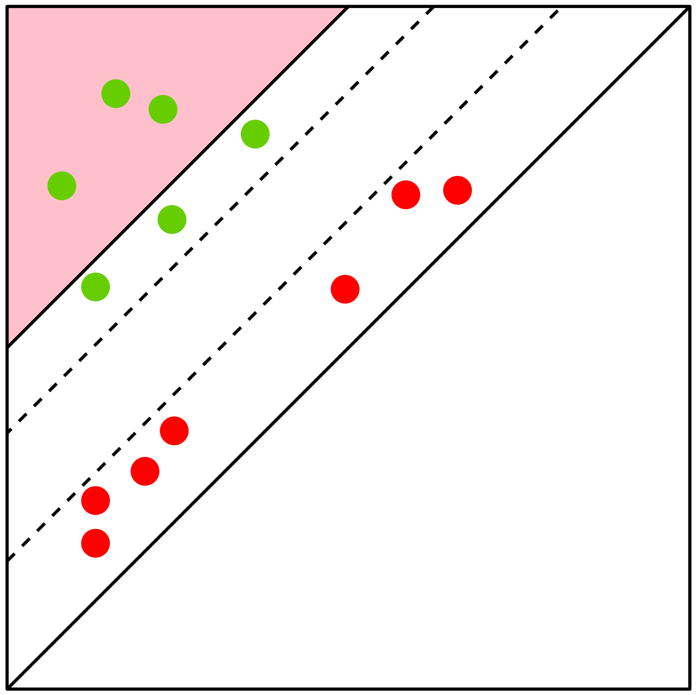
THEOREM[*M., Mukherjee, Harer*] The space D_p is complete and separable.

The proof of completeness of ℓ_p does not translate well here because the notion of a pointwise convergence is not defined.

Instead, we define *persistence-wise* convergence.

The problem is convergence from below.





Can be resolved due to persistence-wise separation of points in the diagrams.

Fréchet Expectation

Definition. Given a probability space $(D_p, \mathcal{B}(D_p), \mathcal{P})$ the quantity

$$\mathrm{Var}_{\mathcal{P}} = \inf_{d \in D_p} \left[F_{\mathcal{P}}(d) = \int_{D_p} W_p(d,e)^2 \, \mathcal{P}(\mathrm{d}e) < \infty \right],$$

is the Fréchet variance of $\mathcal{P}$ and the set at which the value is obtained

$$\mathbb{E}_{\mathcal{P}} = \{ d \mid F_{\mathcal{P}}(d) = \mathrm{Var}_{\mathcal{P}} \},$$

is the Fréchet expectation.

In general, the Fréchet expectation may be empty.

Fréchet Expectation

DEFINITION. Given a probability space $(D_p, \mathcal{B}(D_p), \mathcal{P})$ the quantity

$$\mathrm{Var}_{\mathcal{P}} = \inf_{d \in D_p} \left[F_{\mathcal{P}}(d) = \int_{D_p} W_p(d,e)^2 \, \mathcal{P}(\mathrm{d}e) < \infty \right],$$

is the Fréchet variance of $\mathcal{P}$ and the set at which the value is obtained

$$\mathbb{E}_{\mathcal{P}} = \{ d \mid F_{\mathcal{P}}(d) = \mathrm{Var}_{\mathcal{P}} \},$$

is the Fréchet expectation.

In general, the Fréchet expectation may be empty.

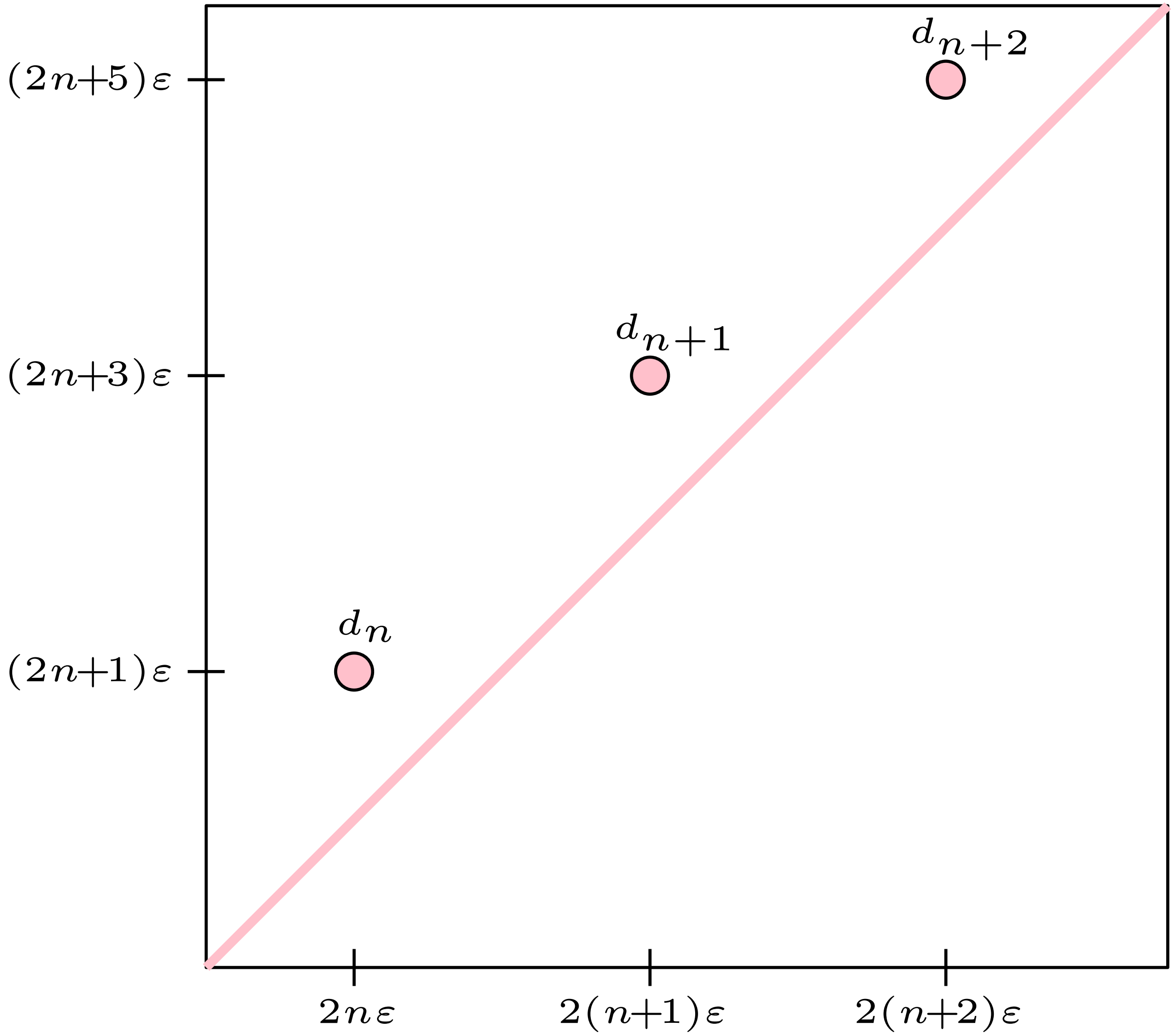
To answer the existence question we need to understand compactness in D_p

Compactness in D_p

Clearly, we need a boundedness. What else can go wrong?

Compactness in D_p

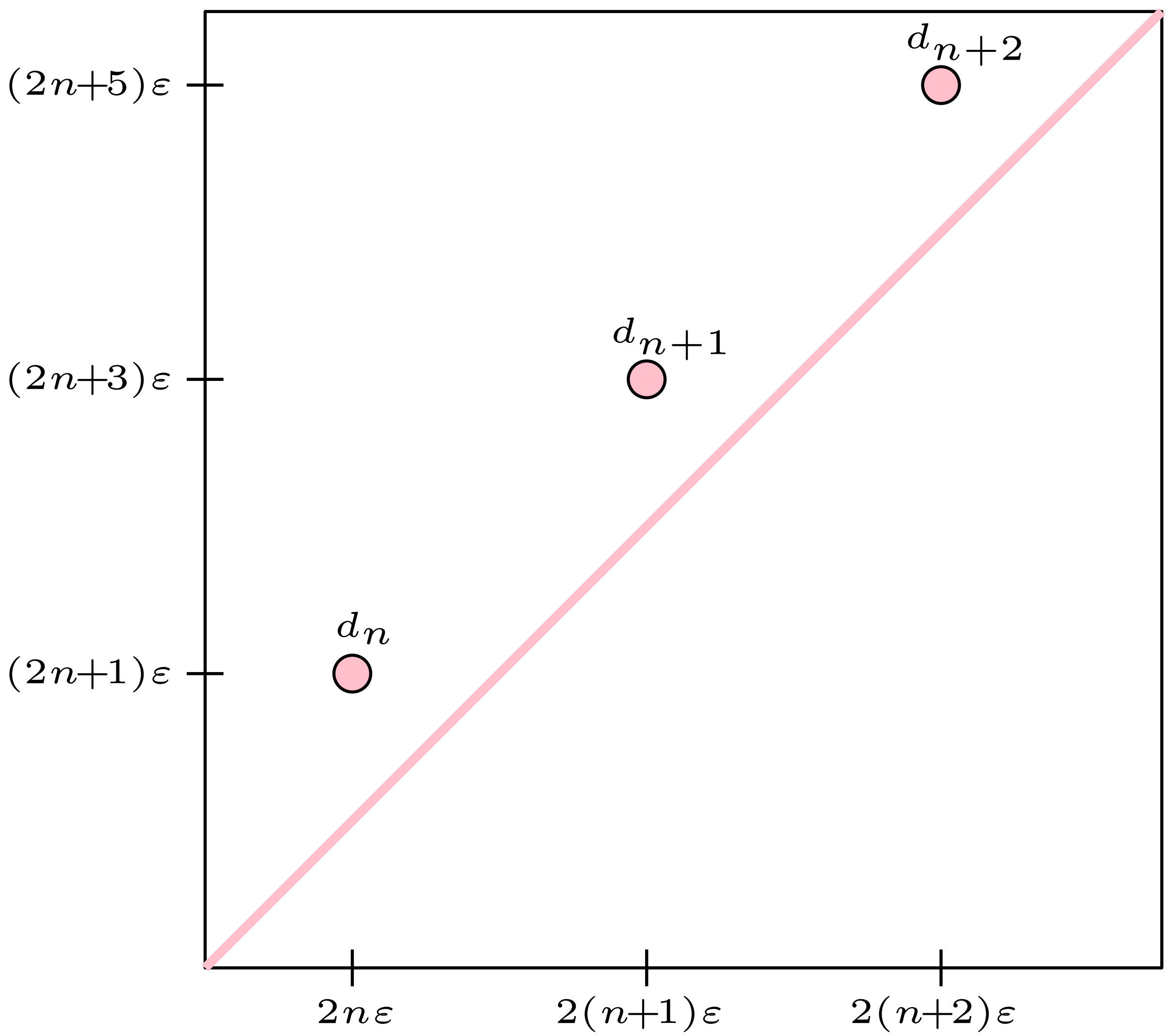
Clearly, we need a boundedness. What else can go wrong?



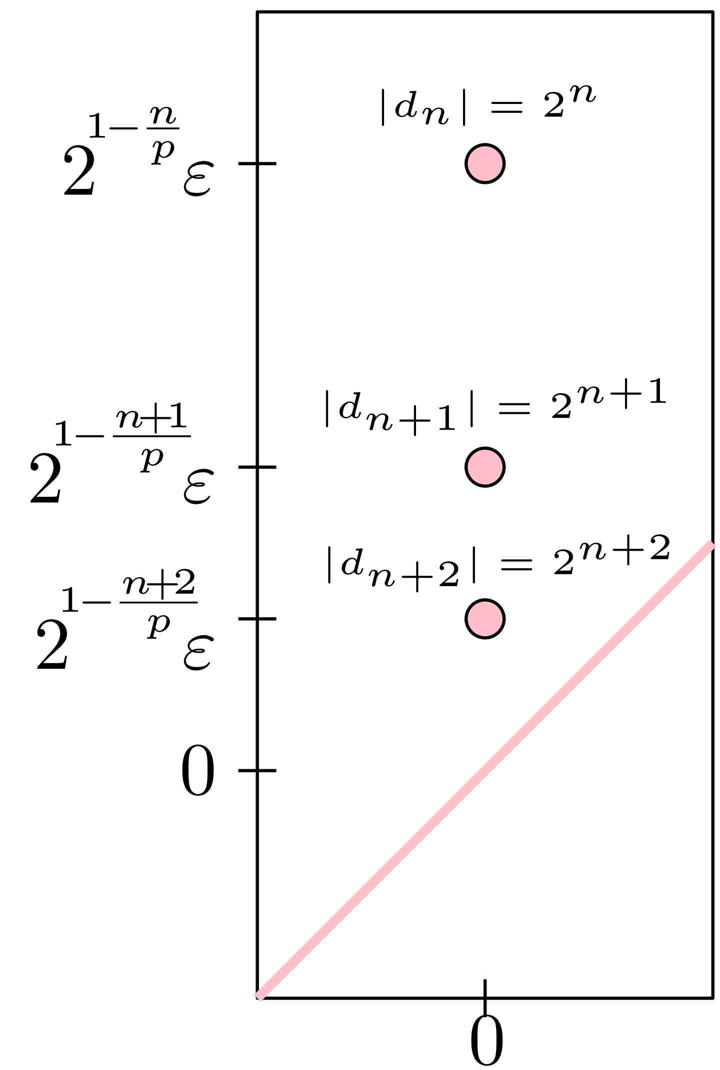
Not off diagonally birth-death bounded

Compactness in D_p

Clearly, we need a boundedness. What else can go wrong?



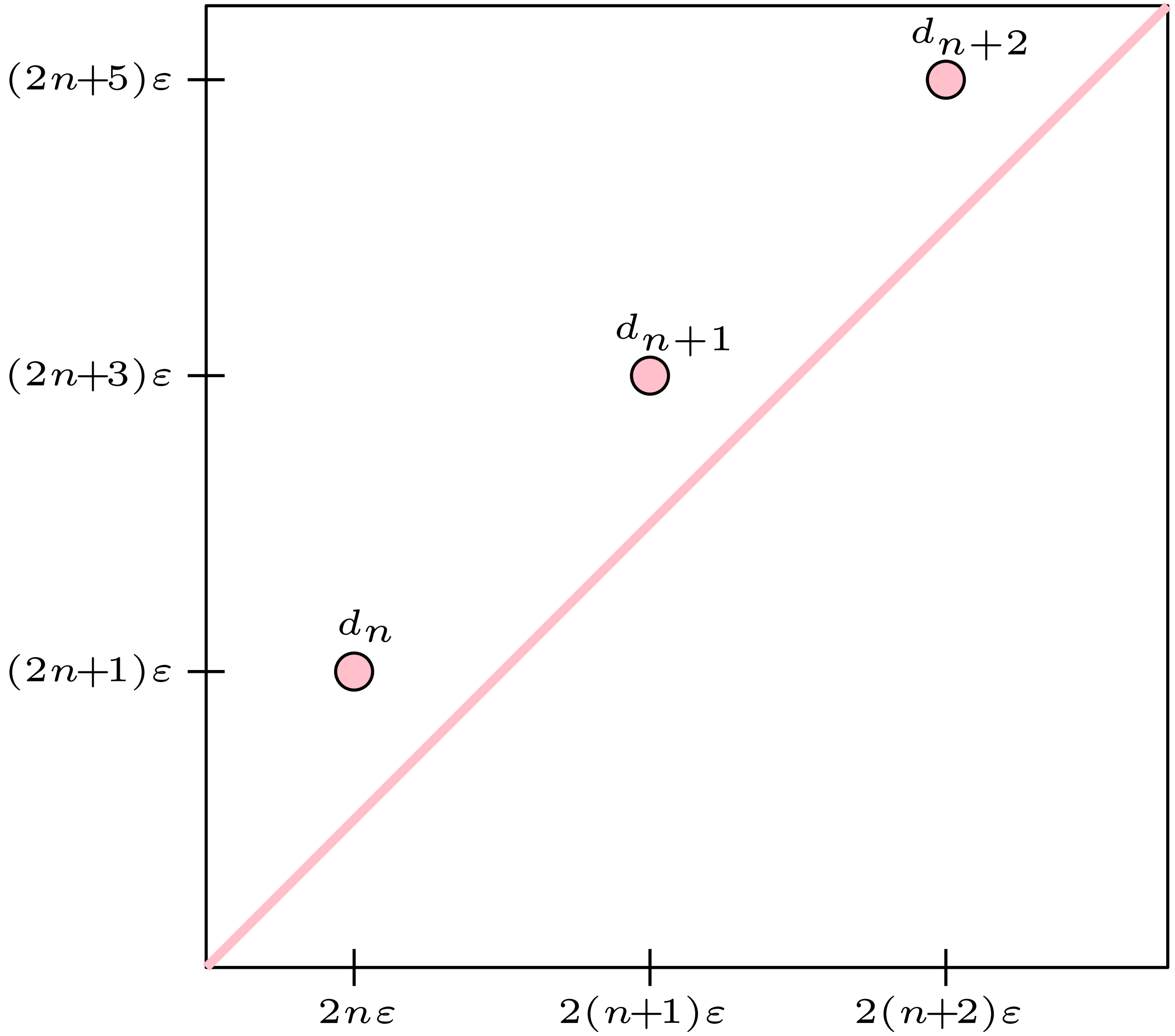
Not off diagonally birth-death bounded



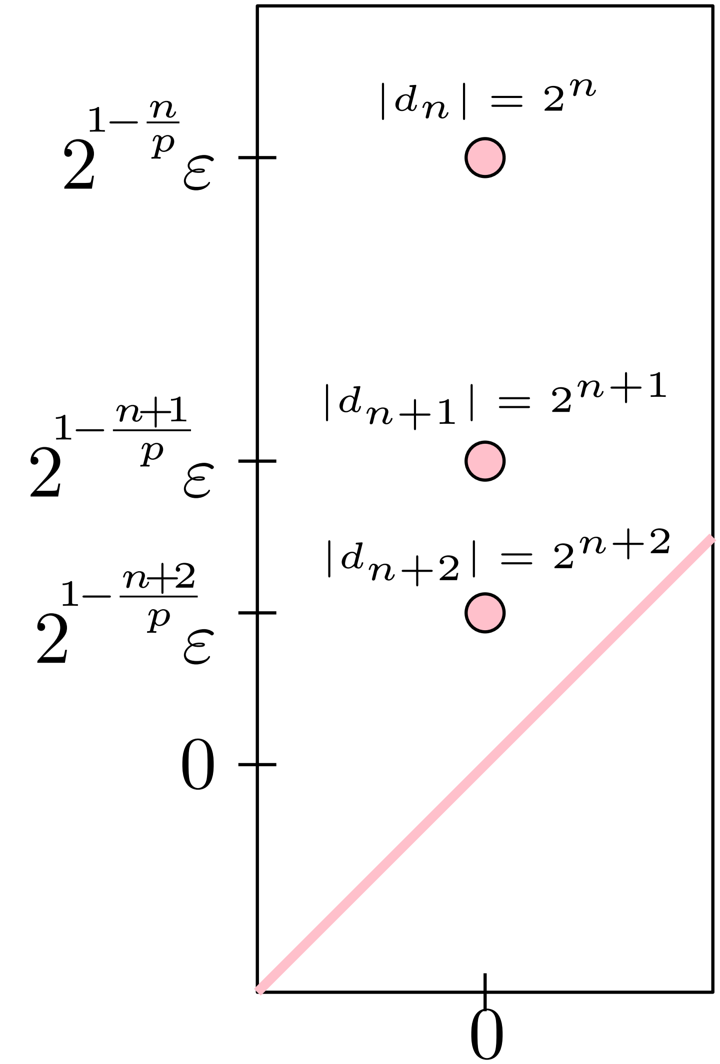
Not uniform

Compactness in D_p

Clearly, we need a boundedness. What else can go wrong?



Not off diagonally birth-death bounded



Not uniform

THEOREM[M., Mukherjee, Harer] A subset of D_p is relatively compact if and only if it is bounded, off-diagonally birth-death bounded, and uniform.

Fréchet Expectation

THEOREM[*Harer, M., Mukherjee*] Let $\mathcal{P}$ be a probability measure on $(D_p, \mathcal{B}(D_p))$ with a finite second moment. If $\mathcal{P}$ has compact support then $\mathbb{E}_{\mathcal{P}} \neq \emptyset$.

Fréchet Expectation

THEOREM[*Harer, M., Mukherjee*] Let $\mathcal{P}$ be a probability measure on $(D_p, \mathcal{B}(D_p))$ with a finite second moment. If $\mathcal{P}$ has compact support then $\mathbb{E}_{\mathcal{P}} \neq \emptyset$.

DEFINITION. A measure μ on a measurable metric space $(\mathbb{X}, \rho)$ is called tight if $\forall \varepsilon > 0$ there exists a compact set $C \subset \mathbb{X}$ such that $\mu(\mathbb{X} - C) < \varepsilon$

DEFINITION. We say that a measure μ on a measurable metric space $(\mathbb{X}, \rho)$ has a rate of decay at infinity q if for some (hance all) $x_0 \in \mathbb{X}$ there exist constants $C > 0$ and $R > 0$ such that for all $r \geq R$ we have $\mu(B^r(x_0)) \leq Cr^{-q}$, where $B^r(x_0) = \{x \in \mathbb{X} | \rho(x, x_0) \geq r\}$.

Fréchet Expectation

THEOREM[*Harer, M., Mukherjee*] Let $\mathcal{P}$ be a probability measure on $(D_p, \mathcal{B}(D_p))$ with a finite second moment. If $\mathcal{P}$ has compact support then $\mathbb{E}_{\mathcal{P}} \neq \emptyset$.

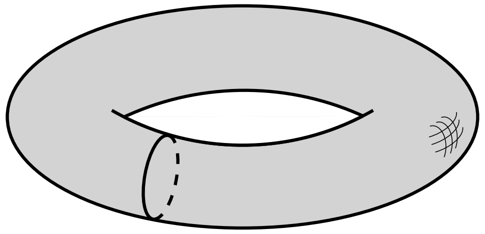
DEFINITION. A measure μ on a measurable metric space $(\mathbb{X}, \rho)$ is called tight if $\forall \varepsilon > 0$ there exists a compact set $C \subset \mathbb{X}$ such that $\mu(\mathbb{X} - C) < \varepsilon$

DEFINITION. We say that a measure μ on a measurable metric space $(\mathbb{X}, \rho)$ has a rate of decay at infinity q if for some (hance all) $x_0 \in \mathbb{X}$ there exist constants $C > 0$ and $R > 0$ such that for all $r \geq R$ we have $\mu(B^r(x_0)) \leq Cr^{-q}$, where $B^r(x_0) = \{x \in \mathbb{X} | \rho(x, x_0) \geq r\}$.

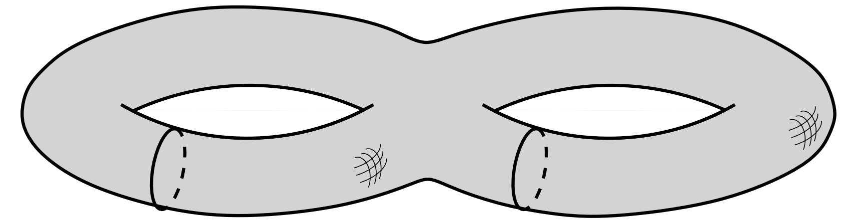
THEOREM[*Harer, M., Mukherjee*] Let $\mathcal{P}$ be a tight probability measure on $(D_p, \mathcal{B}(D_p))$ with the rate of decay at infinity $q > \max\{2, p\}$. Then $\mathbb{E}_{\mathcal{P}} \neq \emptyset$.

Example

Diagrams $\{x_1, \ldots, x_n\} \subset D_p$ of samples of a torus, $\mathcal{O}_1$.

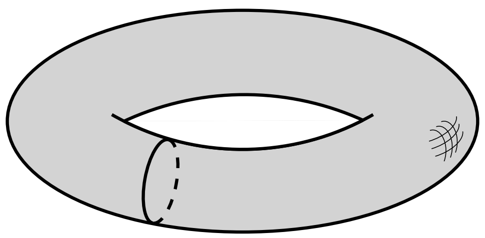


Diagrams $\{y_1, \ldots, y_n\} \subset D_p$ of samples of a double torus, $\mathcal{O}_2$.

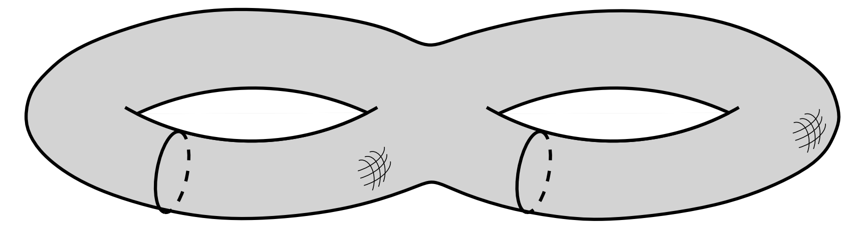


Example

Diagrams $\{x_1, \ldots, x_n\} \subset D_p$ of samples of a torus, $\mathcal{O}_1$.



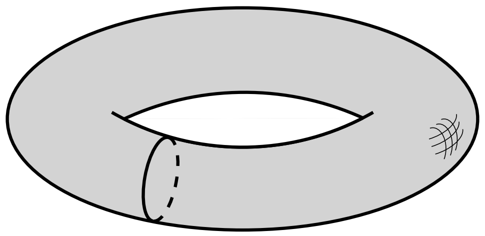
Diagrams $\{y_1, \ldots, y_n\} \subset D_p$ of samples of a double torus, $\mathcal{O}_2$.



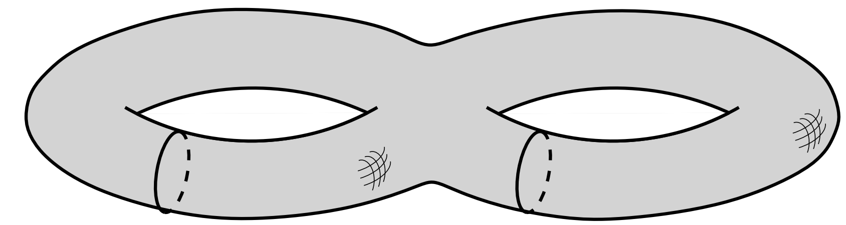
$$\xleftarrow{?} \; z \in D_p \; \xrightarrow{?}$$

Example

Diagrams $\{x_1, \ldots, x_n\} \subset D_p$ of samples of a torus, $\mathcal{O}_1$.



Diagrams $\{y_1, \ldots, y_n\} \subset D_p$ of samples of a double torus, $\mathcal{O}_2$.



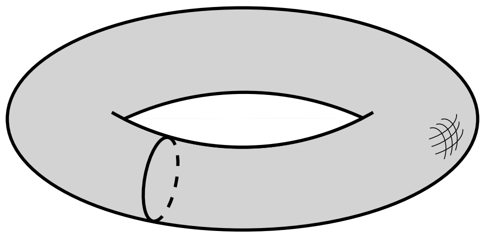
$$\xleftarrow{?} \; z \in D_p \; \xrightarrow{?}$$

$$\hat{p}(x \mid \mathcal{O}_1) = \frac{1}{m\kappa_\tau} \sum_{i=1}^{m} e^{-W_p^2(x, x_i)/\tau},$$

$$\hat{p}(x \mid \mathcal{O}_2) = \frac{1}{n\kappa_\tau} \sum_{i=1}^{n} e^{-W_p^2(x, y_i)/\tau},$$

Example

Diagrams $\{x_1, \ldots, x_n\} \subset D_p$ of samples of a torus, $\mathcal{O}_1$.

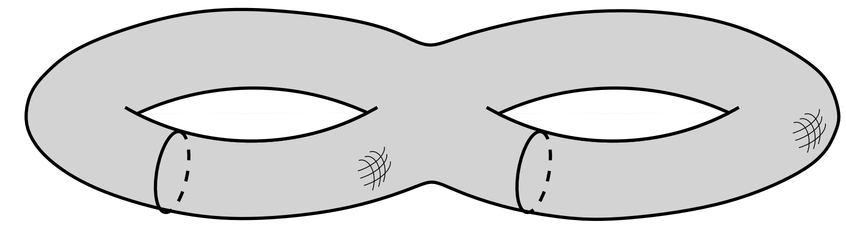


$$\hat{p}(x \mid \mathcal{O}_1) = \frac{1}{m\kappa_\tau} \sum_{i=1}^{m} e^{-W_p^2(x, x_i)/\tau},$$

Prior $\pi_1 = \Pr(\mathcal{O}_1)$

$$\xleftarrow{?} \; z \in D_p \; \xrightarrow{?}$$

Diagrams $\{y_1, \ldots, y_n\} \subset D_p$ of samples of a double torus, $\mathcal{O}_2$.

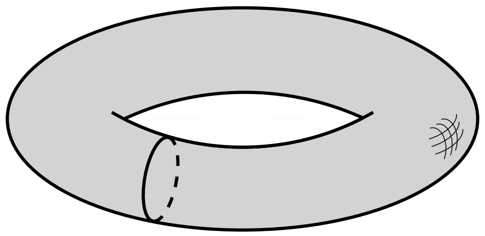


$$\hat{p}(x \mid \mathcal{O}_2) = \frac{1}{n\kappa_\tau} \sum_{i=1}^{n} e^{-W_p^2(x, y_i)/\tau},$$

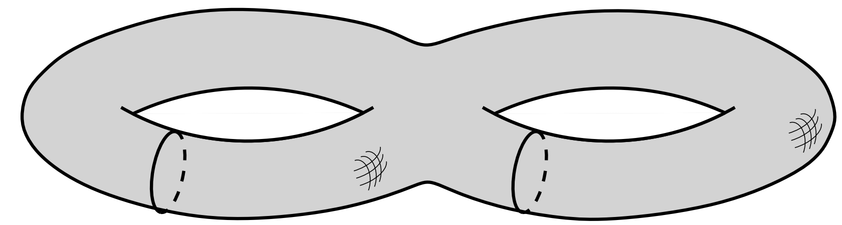
Prior $\pi_2 = \Pr(\mathcal{O}_2)$

Example

Diagrams $\{x_1, \ldots, x_n\} \subset D_p$ of samples of a torus, $\mathcal{O}_1$.



Diagrams $\{y_1, \ldots, y_n\} \subset D_p$ of samples of a double torus, $\mathcal{O}_2$.



$$\xleftarrow{?} \quad z \in D_p \quad \xrightarrow{?}$$

$$\hat{p}(x \mid \mathcal{O}_1) = \frac{1}{m\kappa_\tau} \sum_{i=1}^{m} e^{-W_p^2(x,x_i)/\tau},$$

$$\hat{p}(x \mid \mathcal{O}_2) = \frac{1}{n\kappa_\tau} \sum_{i=1}^{n} e^{-W_p^2(x,y_i)/\tau},$$

Prior $\pi_1 = \Pr(\mathcal{O}_1)$

Prior $\pi_2 = \Pr(\mathcal{O}_2)$

Posterior

$$\hat{p}(\mathcal{O}_1 \mid z) = \frac{\hat{p}(z \mid \mathcal{O}_1)\pi_1}{\hat{p}(z)} = \frac{\hat{p}(z \mid \mathcal{O}_1)\pi_1}{\hat{p}(z \mid \mathcal{O}_1)\pi_1 + \hat{p}(z \mid \mathcal{O}_2)\pi_2}$$

Future Directions

- Algorithm for computing Fréchet mean.
  - $\exists$ algorithm for computing Fréchet mean in NPC spaces.
  - D_p is not NPC and the Fréchet mean is not unique in D_p.

Future Directions

- Algorithm for computing Fréchet mean.
  - $\exists$ algorithm for computing Fréchet mean in NPC spaces.
  - D_p is not NPC and the Fréchet mean is not unique in D_p.

- Convergence results
  - We would like to show that persistence diagrams computed for samples from a probability measure converge to the persistence diagram of the actual probability measure.

Future Directions

- Algorithm for computing Fréchet mean.
  - $\exists$ algorithm for computing Fréchet mean in NPC spaces.
  - D_p is not NPC and the Fréchet mean is not unique in D_p.

- Convergence results
  - We would like to show that persistence diagrams computed for samples from a probability measure converge to the persistence diagram of the actual probability measure.

- Similar results for other topological summaries.
  - For example, Reeb graphs.

Collaborators

John Harer

Sayan Mukherjee

Funding

References

Y. Mileyko, S. Mukherjee, and J. Harer. Probability measures on the space of persistence diagrams. Inverse Problems, in press.

D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko. Lipschitz functions have L_p-stable persistence, Foundations of Computational Mathematics Journal, Vol.10, No.2, pp. 127–139, 2010

H. Edelsbrunner, D. Letscher and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.* 28 (2002), 511–533.

Thank you.