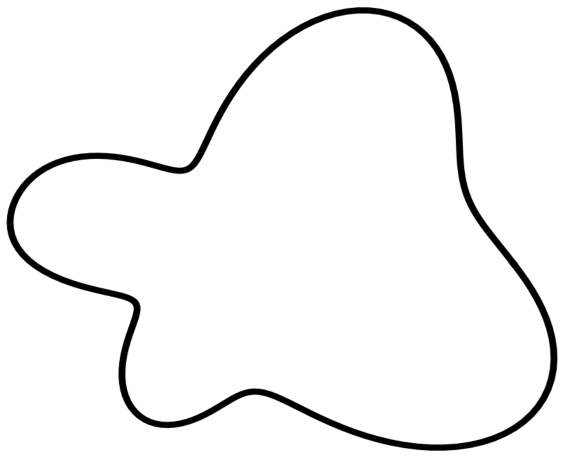
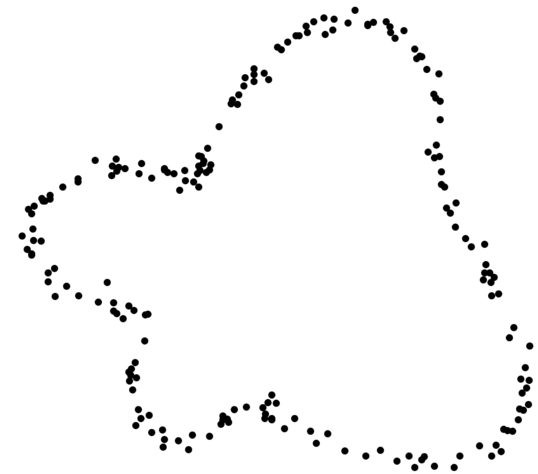
DTM-based filtrations

Hirokazu Anai, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hiroya Inakoshi, Raphaël Tinarrage, Yuhei Umeda

Usual pipeline of persistent homology

Let $M \subset \mathbb{R}^n$ be a compact set, and $X \subset \mathbb{R}^n$ a finite set.

M 

 X

Denote by $H_i(M)$ the i^{th} singular homology group of M in a finite field k.

$$H_0(M) = k \qquad H_1(M) = k \qquad H_i(M) = 0,\ i \geq 2$$

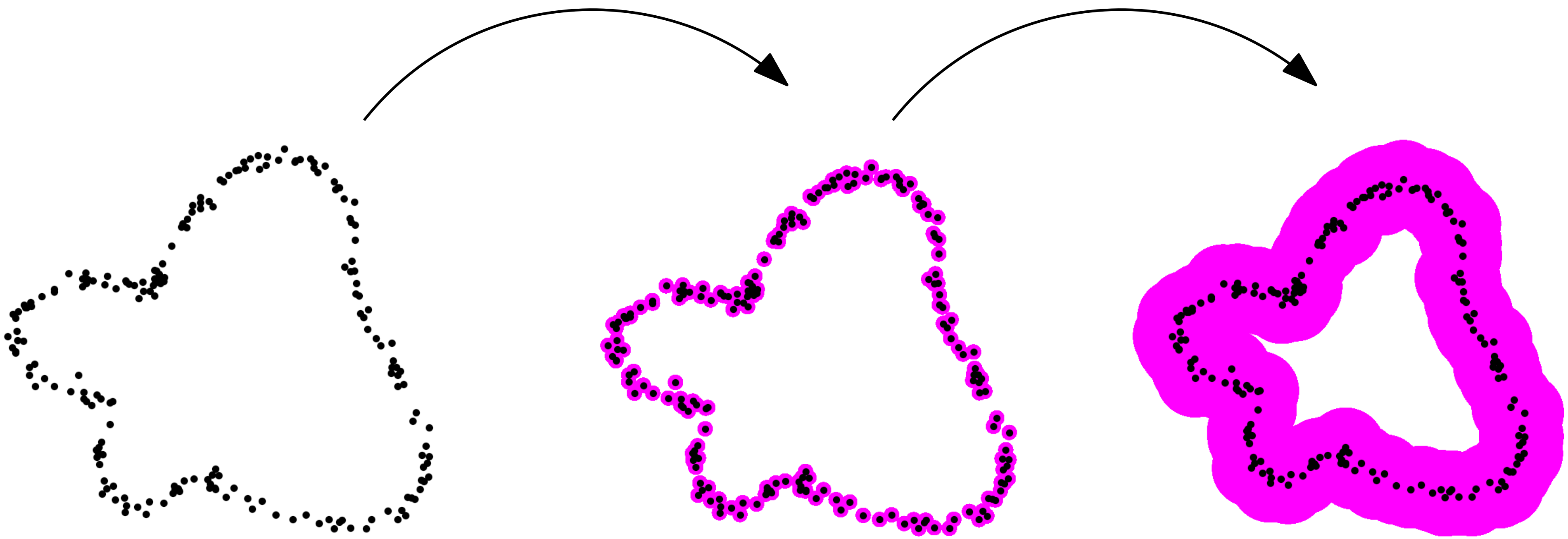
Problem: estimate $H_i(M)$ from X.

Assumption: the Hausdorff distance $d_H(M, X)$ is small.

Čech filtration

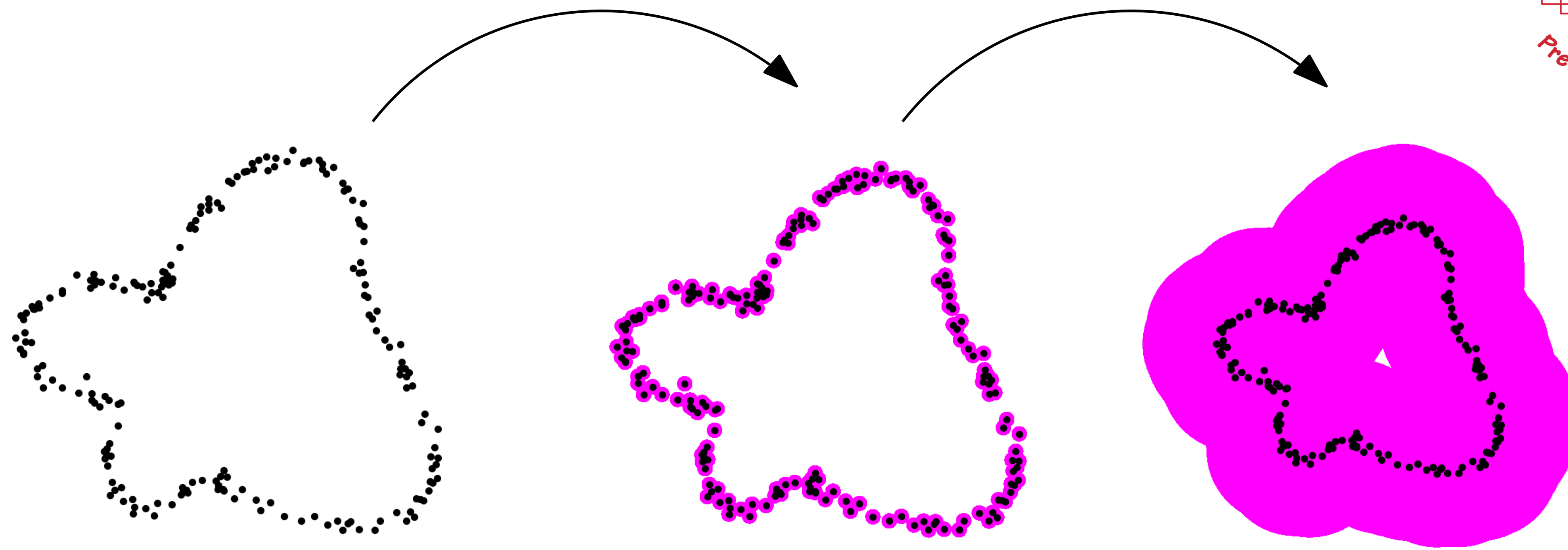
For every $t \geq 0$, define $X^t = \bigcup_{x \in X} \overline{B}(x,t)$, the t-neighborhood of X.

The family $(X^t)_{t \geq 0}$ is called the Čech filtration associated to X.
We have inclusion maps $i_s^t : X^s \to X^t$ for every $0 \leq s \leq t$.



X^t, $t = 0$ $\quad$ $t = 0.05$ $\quad$ $t = 0.2$

Persistence module of the Čech filtration



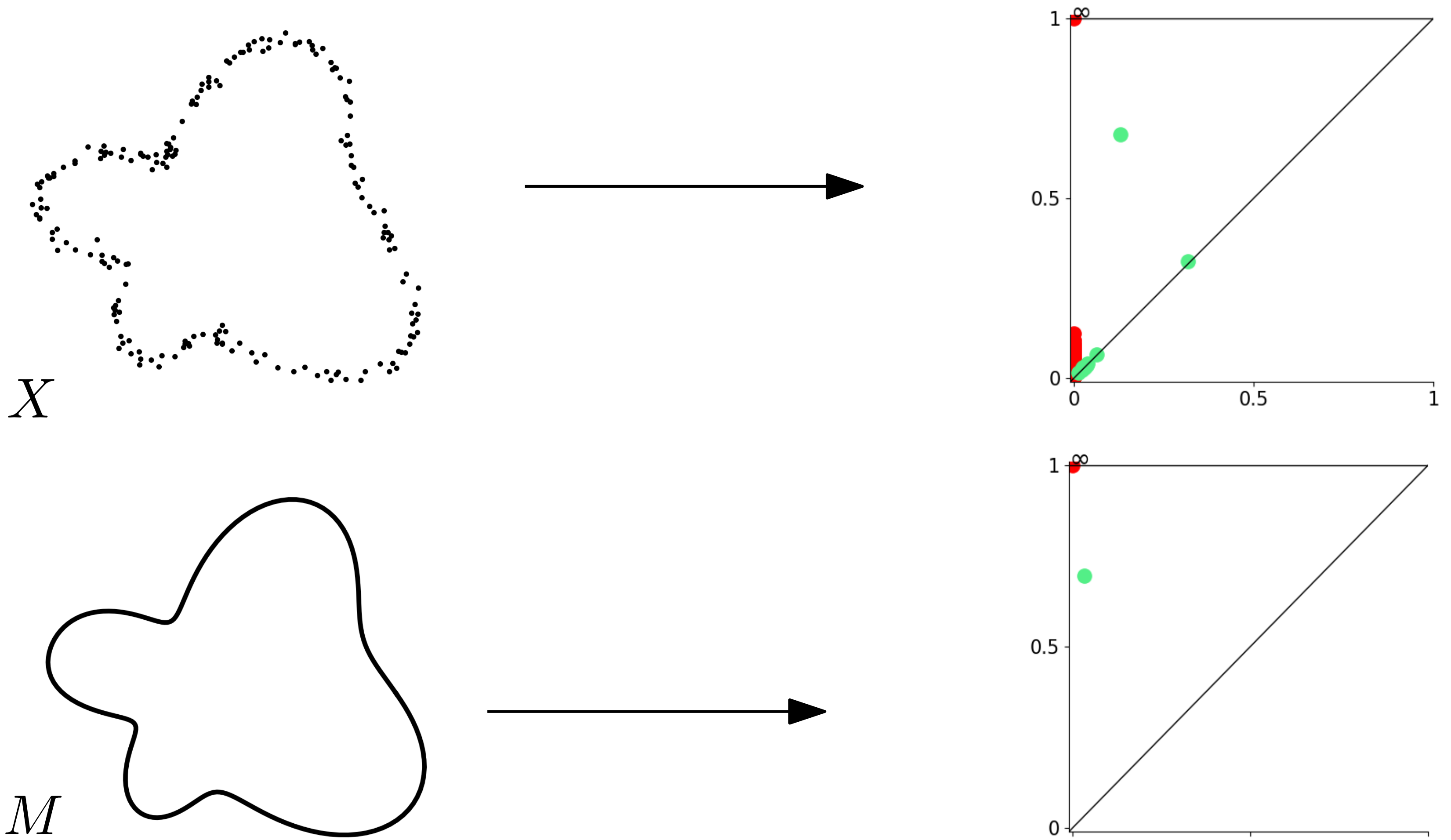
$t = 0$ $\qquad$ $t = 0.05$ $\qquad$ $t = 0.8$

$$
\begin{tikzcd}
H_i(X^t) \arrow[r, bend right] & H_i(X^t) \arrow[r, bend right] & H_i(X^t)
\end{tikzcd}
$$

Applying the homology functor H_i we obtain a persistence module $((H_i(X^t))_{t\geq 0}, ((i_s^t)_*)_{t\geq s\geq 0})$.

Persistence diagram

This persistent module can be summarized via its persistence diagram.

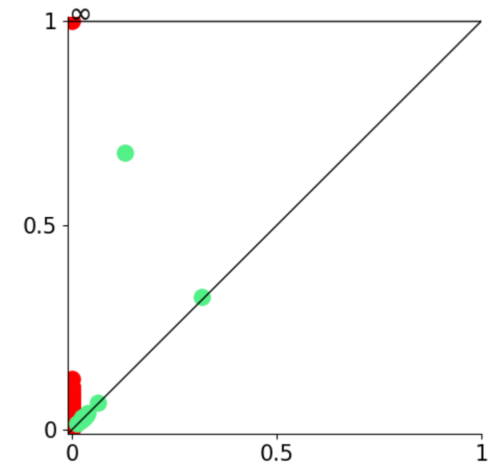


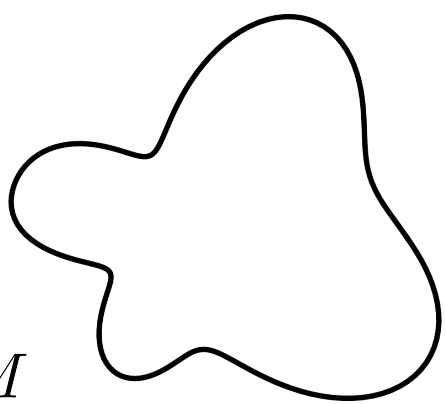
Persistence diagram

Theorem (Stability of persistent homology): The bottleneck distance between these persistence diagrams is upper bounded by the Hausdorff distance $d_H(X, M)$.

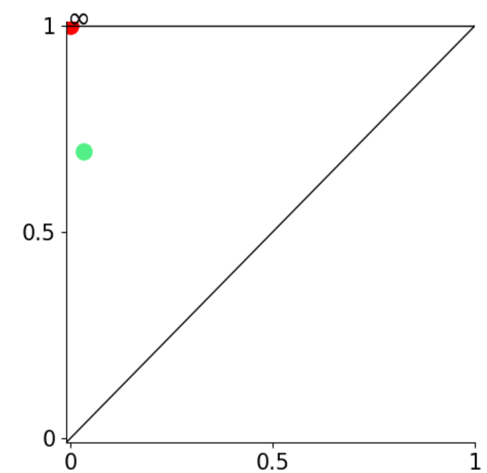
This also holds for the interleaving distance between the (set) filtrations and the interleaving distance between the persistence modules.

X



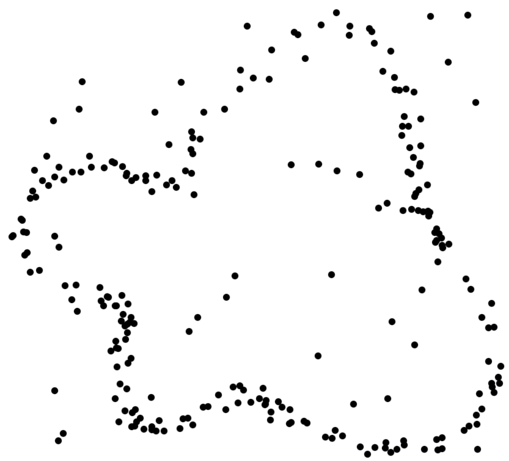


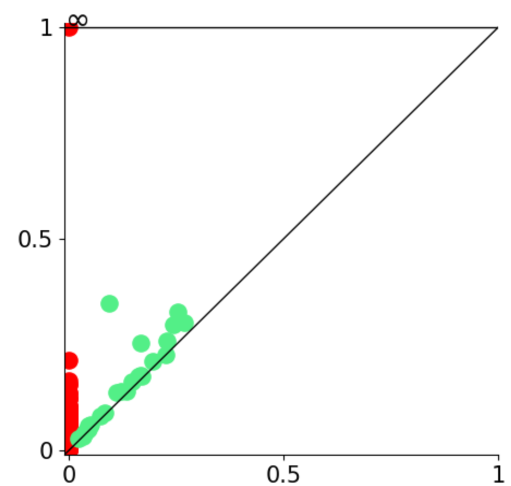
M



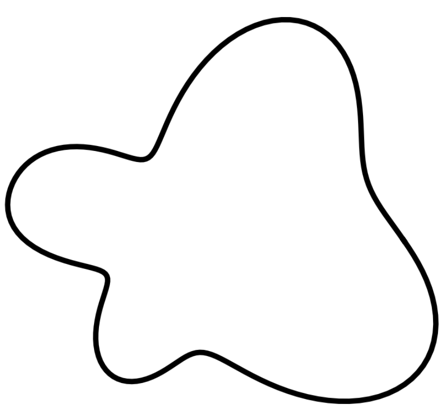
Persistence diagram

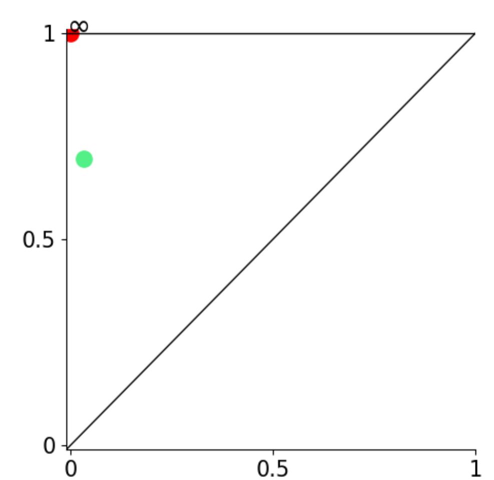
Issue: if X contains outliers, $d_H(M, X)$ is large, and so may be the interleaving and bottleneck distance.





X, a sample of M with outliers





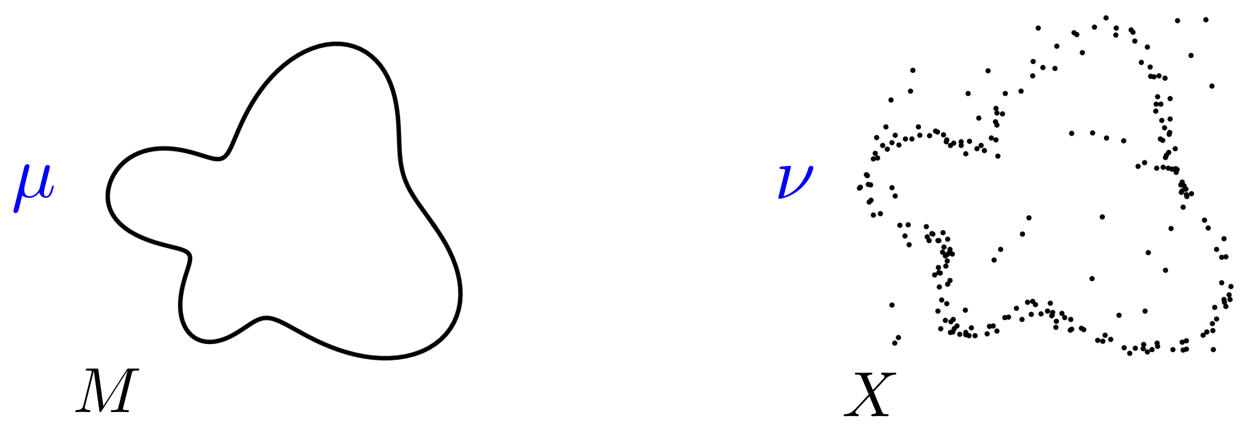
A measure-theoretic point of view

However, X still contains information about M.

Define:

- μ the Hausdorff measure restricted to M (submanifold),
- $\nu = \frac{1}{|X|} \sum_{x \in X} \delta_x$ the empirical measure on X.

$\longrightarrow$ The Wasserstein distance $W_2(\mu, \nu)$ is small.



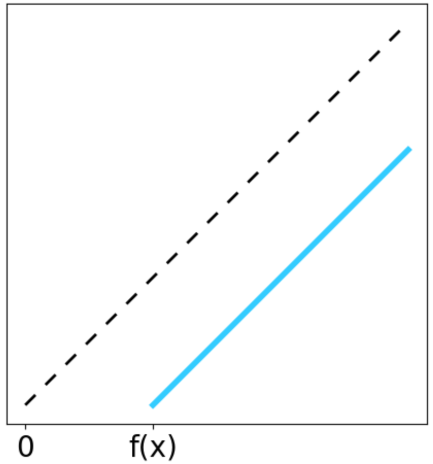
Goal: build filtrations from probability measures which are stable with respect to W_2.

Weighted Čech filtrations

Let $X \subset \mathbb{R}^n$ and $f : X \to \mathbb{R}^+$ (called the weight function).

For every $t \geq 0$, define $V^t[X, f] = \bigcup_{x \in X} \overline{B}(x, t - f(x))$

The family $V[X, f] = (V^t[X, f])_{t \geq 0}$ is called the weighted Čech filtration.



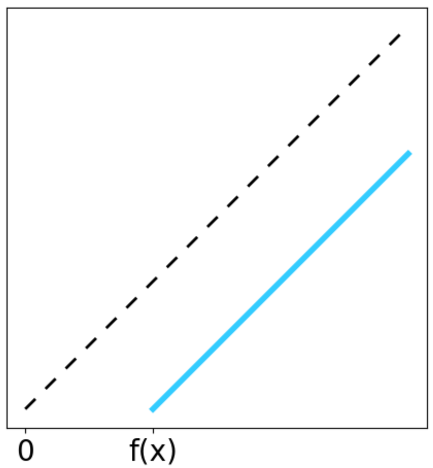
$t \mapsto t - f(x)$

Weighted Čech filtrations

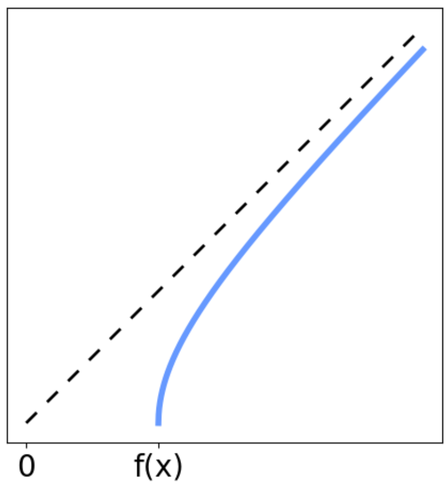
Let $X \subset \mathbb{R}^n$ and $f : X \to \mathbb{R}^+$ (called the weight function).

For every $t \geq 0$, define $V^t[X, f] = \bigcup_{x \in X} \overline{B}(x, \sqrt{t^2 - f(x)^2})$

The family $V[X, f] = (V^t[X, f])_{t \geq 0}$ is called the weighted Čech filtration.



$t \mapsto t - f(x)$



$t \mapsto \sqrt{t^2 - f(x)^2}$

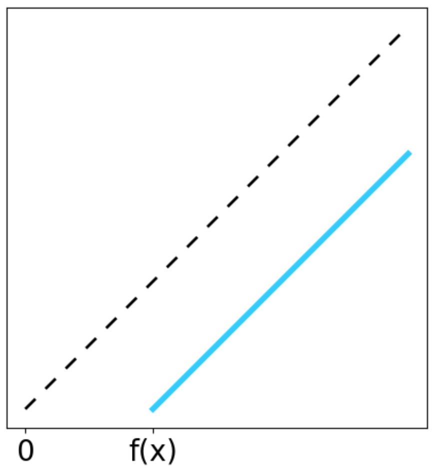
[Buchet et al. SODA 2015]

Weighted Čech filtrations

Let $X \subset \mathbb{R}^n$ and $f : X \to \mathbb{R}^+$ (called the weight function).
Let $p \geq 1$.

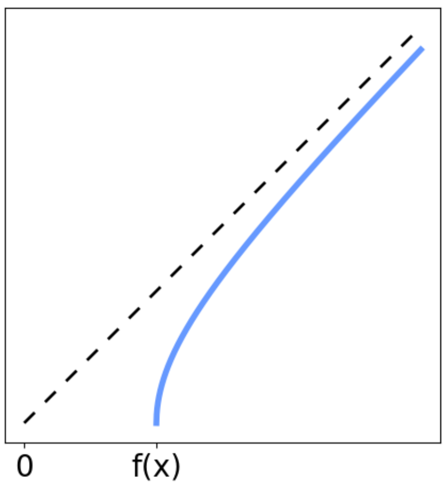
For every $t \geq 0$, define $V^t[X, f, p] = \bigcup_{x \in X} \overline{B}(x, (t^p - f(x)^p)^{\frac{1}{p}})$

The family $V[X, f, p] = (V^t[X, f, p])_{t \geq 0}$ is called the weighted Čech filtration.



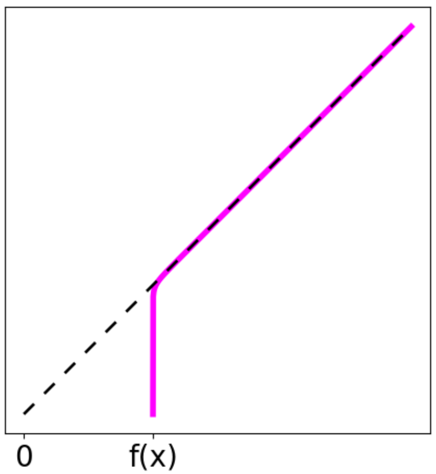
$t \mapsto t - f(x)$

$p = 1$



$t \mapsto \sqrt{t^2 - f(x)^2}$

$p = 2$



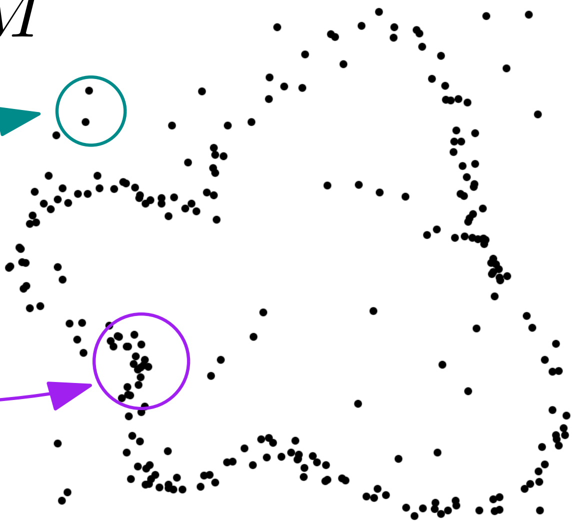
$t \mapsto (t^p - f(x)^p)^{\frac{1}{p}}$

$p = 30$

[Buchet et al. SODA 2015]

How to choose the function f

We want f to take high values far from M
i.e. high on low density areas



And small values close to M
i.e. on small on high density areas

$\longrightarrow$ Choose the distance to measure $d_{\mu,m}$

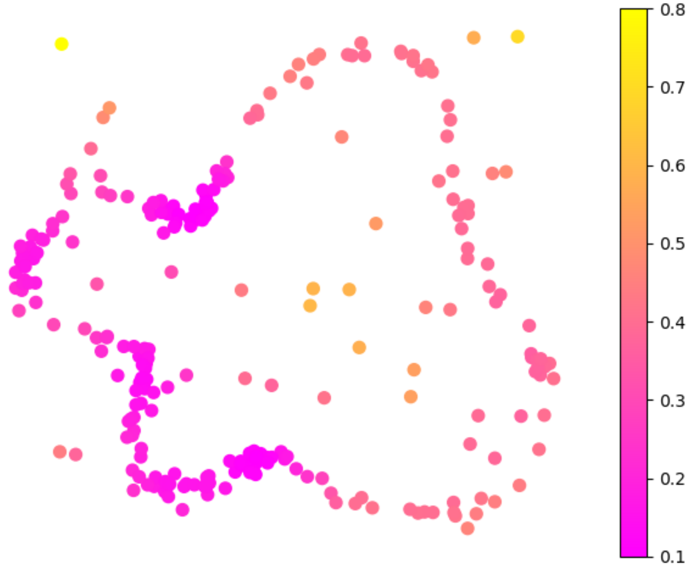
How to choose the function f

Definition (DTM): Let μ be a probability measure.
For every $x \in \mathbb{R}^n$ and $t \in [0,1)$, define

$$\delta_{\mu,t}(x) = \inf\{r \geq 0, \mu(\overline{B}(x,r)) > t\}.$$

Let $m \in [0,1[$. The DTM μ of parameter m is the function:

$$\begin{array}{rccc} d_{\mu,m} : & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \sqrt{\frac{1}{m}\int_0^m \delta_{\mu,t}^2(x)dt} \end{array}$$

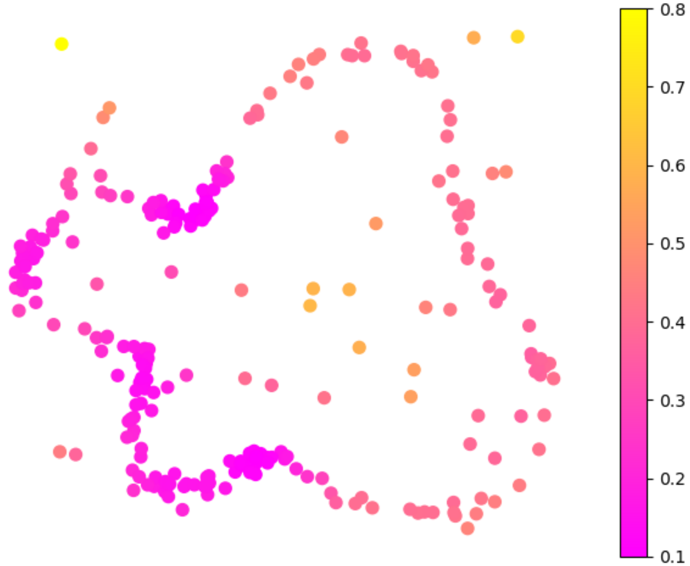


$d_{\mu,m}$, $m = 0.1$

How to choose the function f

Theorem [Chazal et al. 2014]: Let μ, ν be two probability measures, and $m \in (0,1)$. Then

$$\|d_{\mu,m} - d_{\nu,m}\|_\infty \le m^{-\frac{1}{2}} W_2(\mu,\nu).$$



$d_{\mu,m}$, $m = 0.1$

DTM-filtration

Definition: Let μ be a probability measure, $m \in [0,1)$ and $p \geq 1$.
the DTM-filtration is the weighted Čech filtration $V[X, f, p]$ with:

- $X = \mathrm{supp}(\mu)$
- $f = d_{\mu,m}$

It is denoted $W[\mu, m, p]$.

$W[\mu, m, p] = (W^t[\mu, m, p])_{t \geq 0}$ with:

$$W^t[\mu, m, p] = \bigcup_{x \in \mathrm{supp}(\mu)} \overline{B}(x, (t^p - d_{\mu,m}(x)^p)^{\frac{1}{p}})$$

$\mu =$ empirical measure on X
$m = 0.1$

Example

$t = 0, 0$

Usual Čech
$\cup \overline{B}(x, t)$

DTM-filtration, $p = 1$
$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$

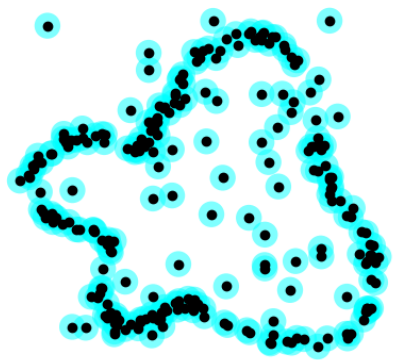
DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

$\mu =$ empirical measure on X
$m = 0.1$

Example

$$t = 0, 1$$



Usual Čech
$\cup \overline{B}(x, t)$

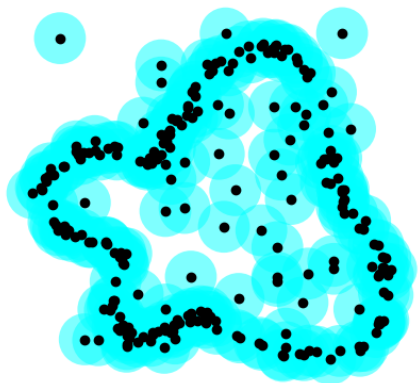
DTM-filtration, $p = 1$
$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$

DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

$\mu =$ empirical measure on X
$m = 0.1$

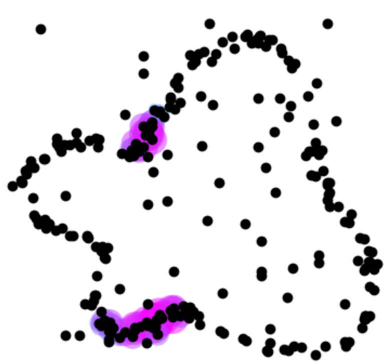
Example

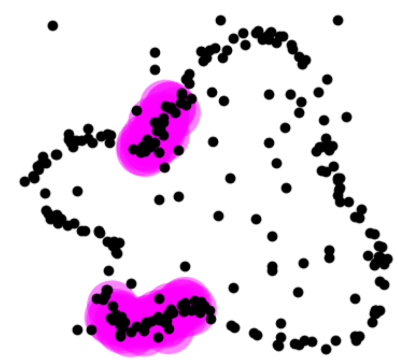


$t = 0, 2$

Usual Čech
$\cup \overline{B}(x, t)$

DTM-filtration, $p = 1$
$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$





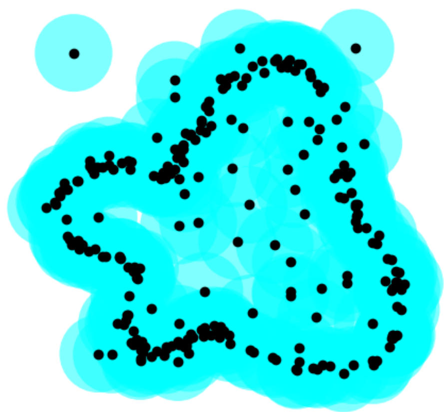
DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

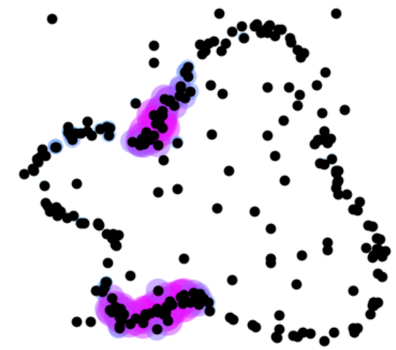
$\mu =$ empirical measure on X
$m = 0.1$

Example

$t = 0,3$

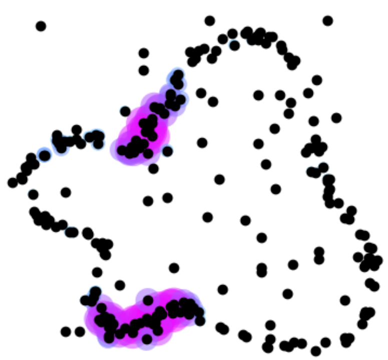


Usual Čech
$\cup \overline{B}(x, t)$

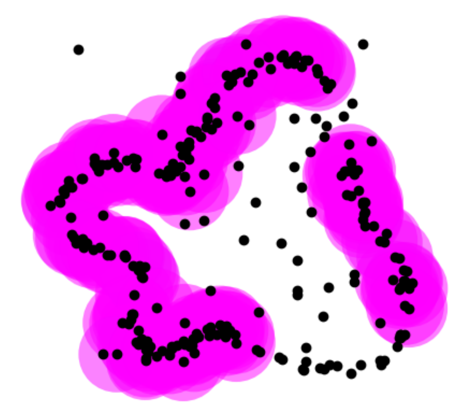


DTM-filtration, $p = 1$

$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$



DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

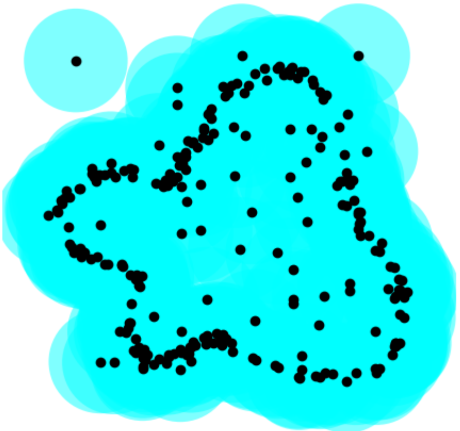


DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

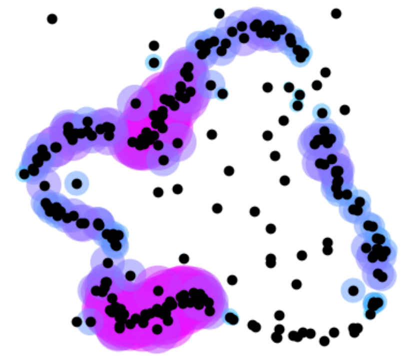
$\mu =$ empirical measure on X
$m = 0.1$

Example

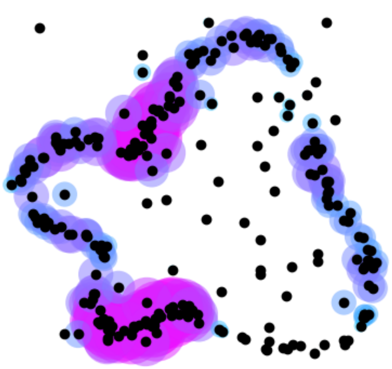
$t = 0,4$



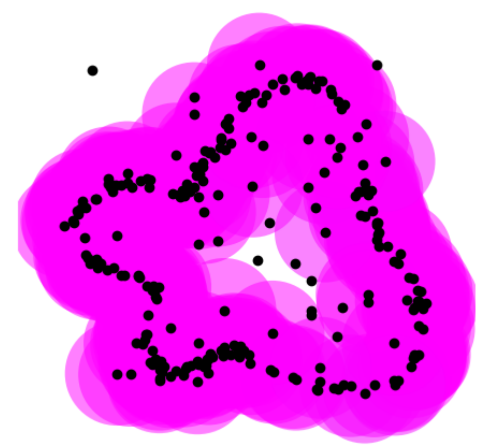
Usual Čech
$\cup\overline{B}(x,t)$



DTM-filtration, $p = 1$
$W^t[\mu, m, p] = \cup\overline{B}(x, t - d_{\mu,m}(x))$



DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup\overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

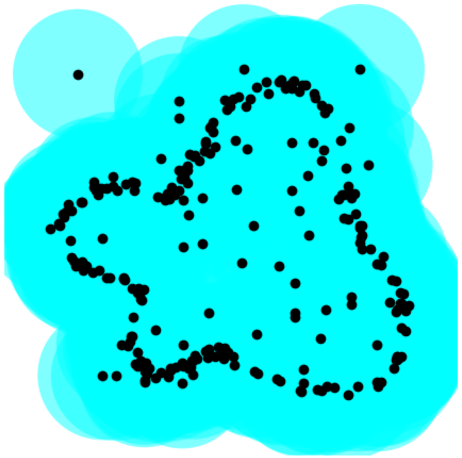


DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup\overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

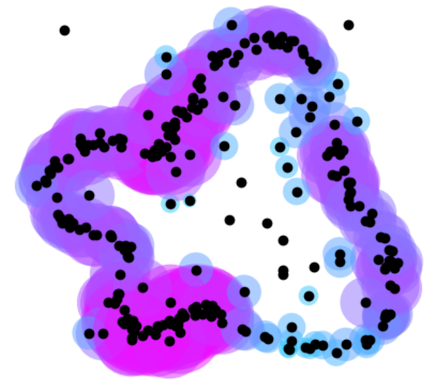
$\mu =$ empirical measure on X
$m = 0.1$

Example

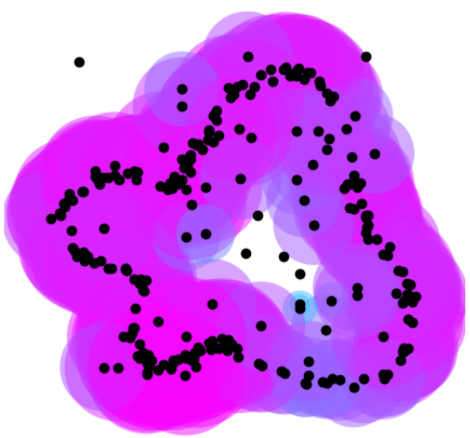
$t = 0,5$



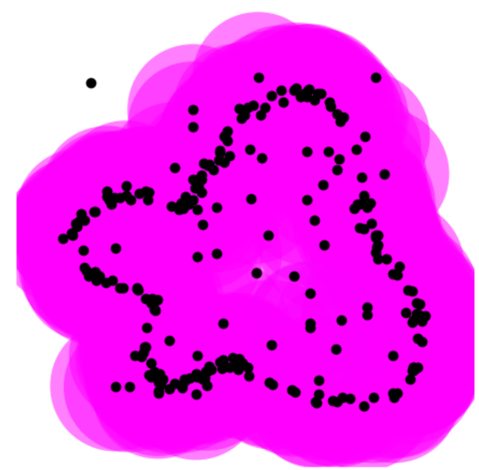
Usual Čech
$\cup \overline{B}(x,t)$



DTM-filtration, $p = 1$
$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$



DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

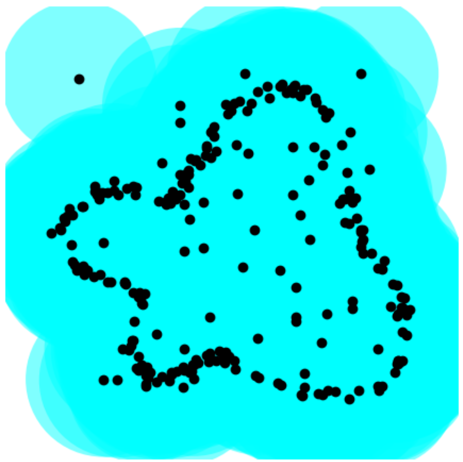


DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

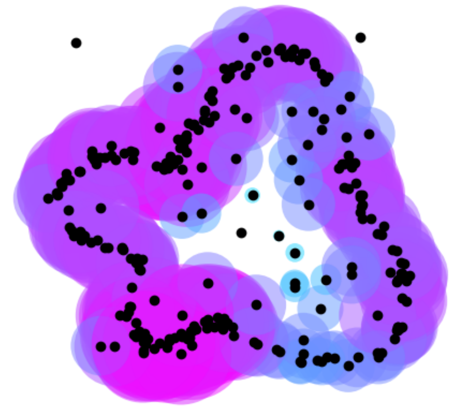
$\mu =$ empirical measure on X
$m = 0.1$

Example

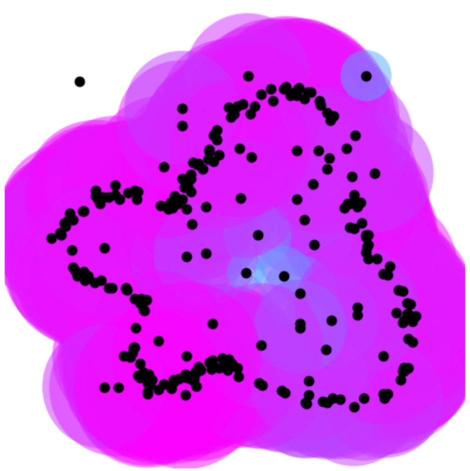
$t = 0, 6$



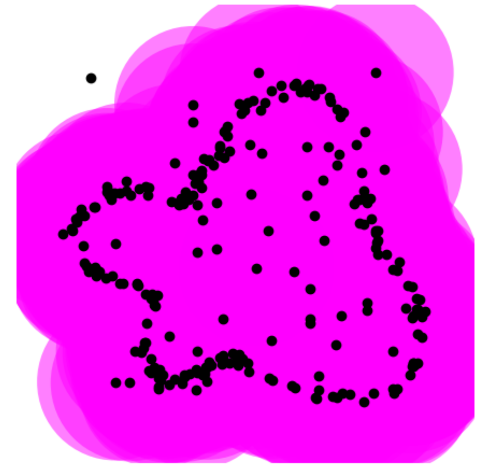
Usual Čech
$\cup \overline{B}(x, t)$



DTM-filtration, $p = 1$
$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$



DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

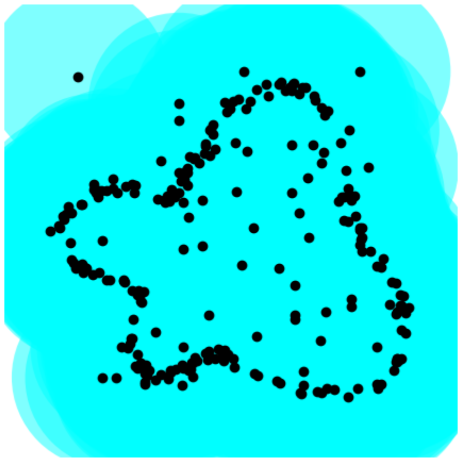


DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

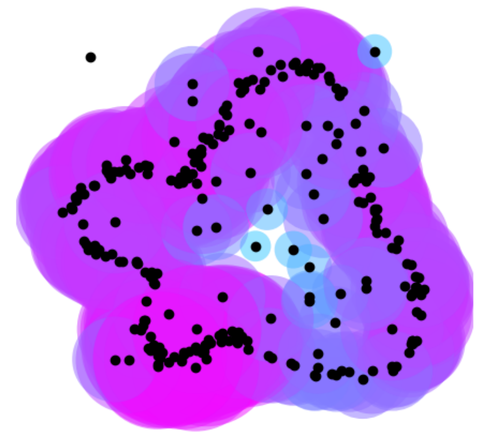
$\mu =$ empirical measure on X
$m = 0.1$

Example

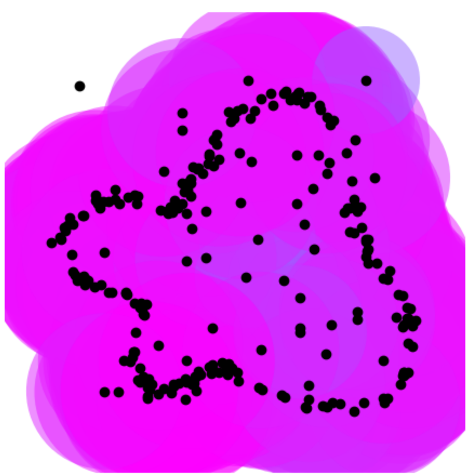
$t = 0,7$



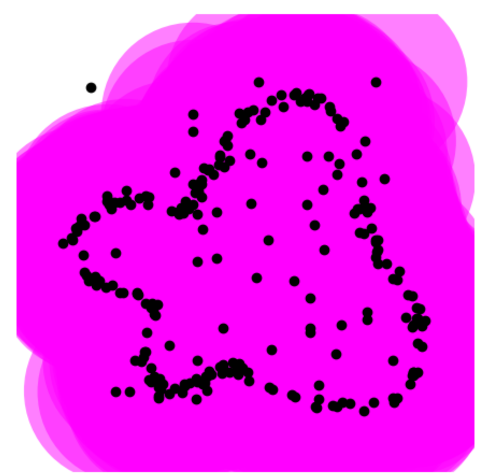
Usual Čech
$\cup \overline{B}(x,t)$



DTM-filtration, $p = 1$
$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$



DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

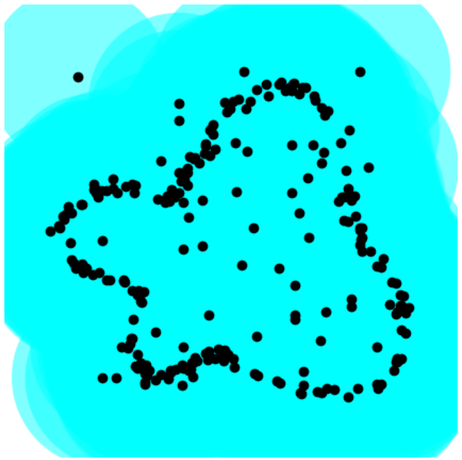


DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

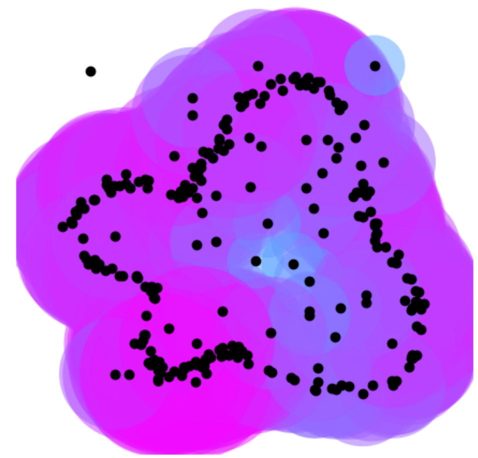
$\mu =$ empirical measure on X
$m = 0.1$

Example

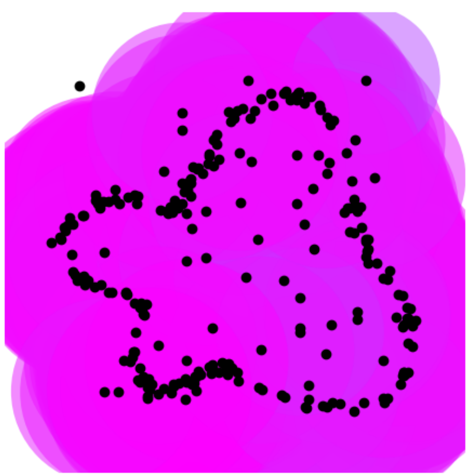
$t = 0,8$



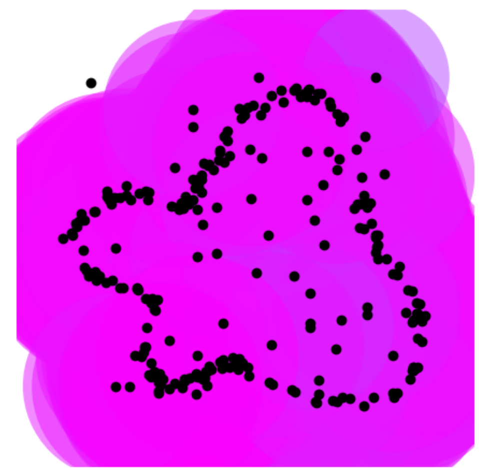
Usual Čech
$\cup \overline{B}(x,t)$



DTM-filtration, $p=1$
$W^t[\mu,m,p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$



DTM-filtration, $p=2$
$W^t[\mu,m,p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

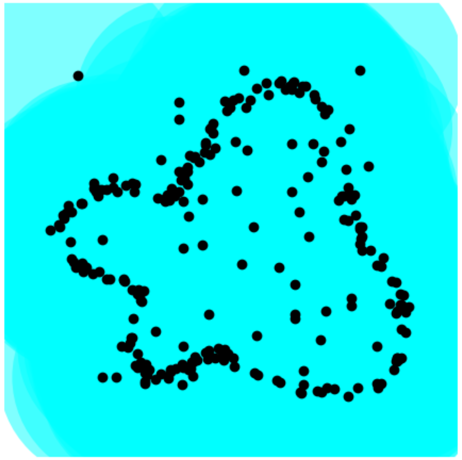


DTM-filtration, $p=30$
$W^t[\mu,m,p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

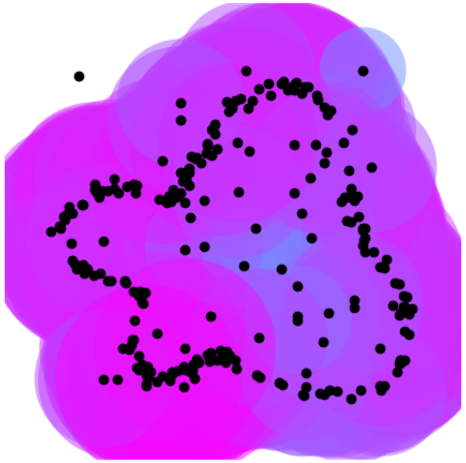
$\mu =$ empirical measure on X
$m = 0.1$

Example

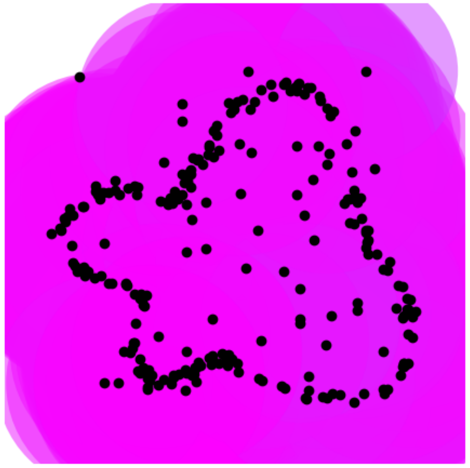
$t = 0,9$



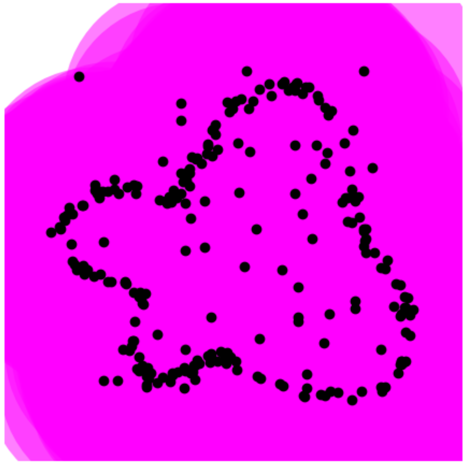
Usual Čech
$\cup \overline{B}(x,t)$



DTM-filtration, $p = 1$
$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$



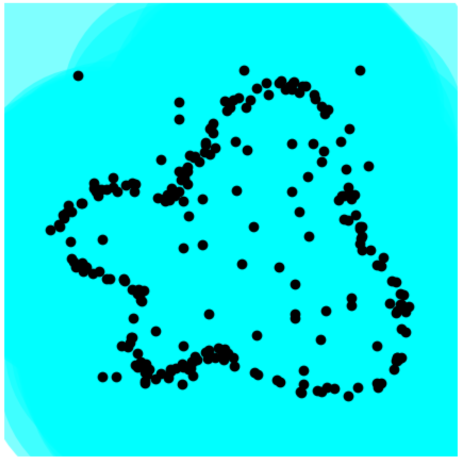
DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$



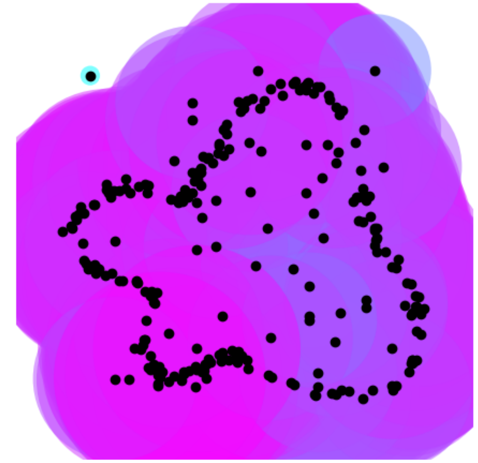
DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

Example

$\mu =$ empirical measure on X
$m = 0.1$

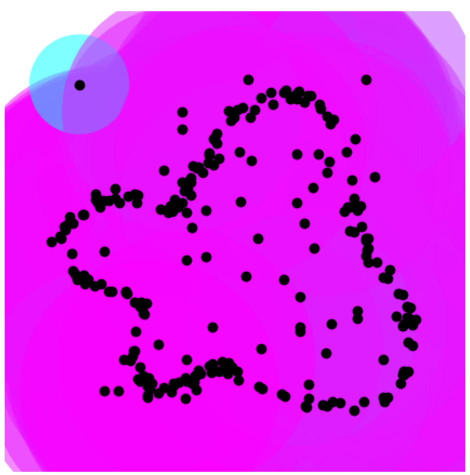


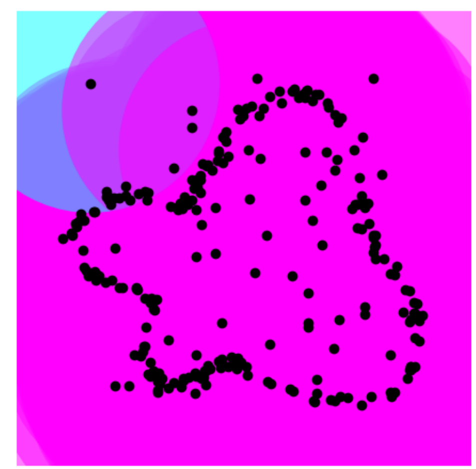
$t = 1, 0$



Usual Čech
$\cup \overline{B}(x, t)$

DTM-filtration, $p = 1$
$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$



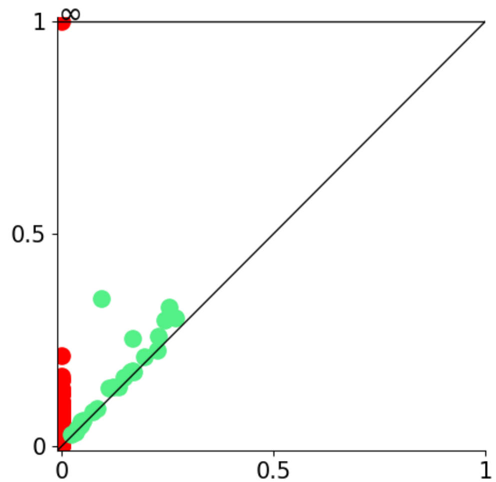


DTM-filtration, $p = 2$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$

DTM-filtration, $p = 30$
$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

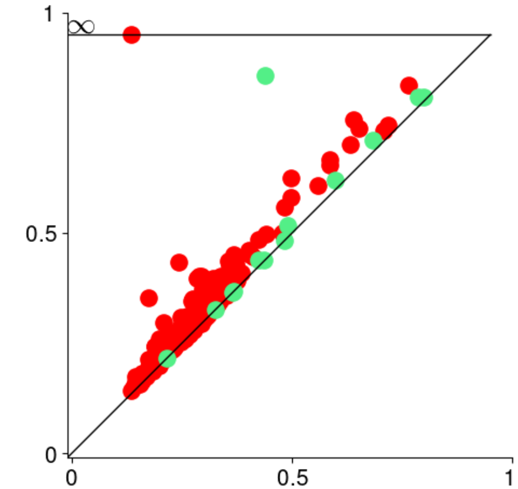
$\mu =$ empirical measure on X
$m = 0.1$

Example



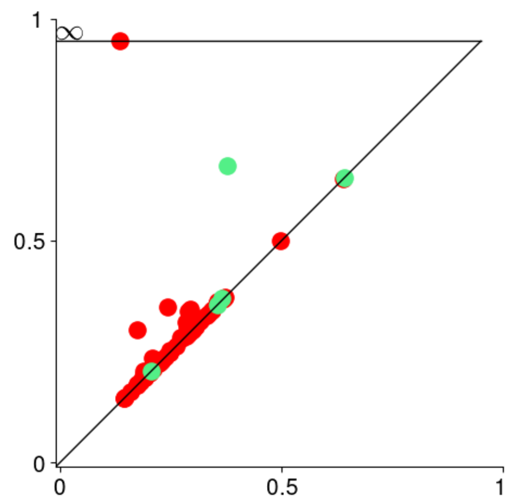
Usual Čech

$\cup \overline{B}(x,t)$



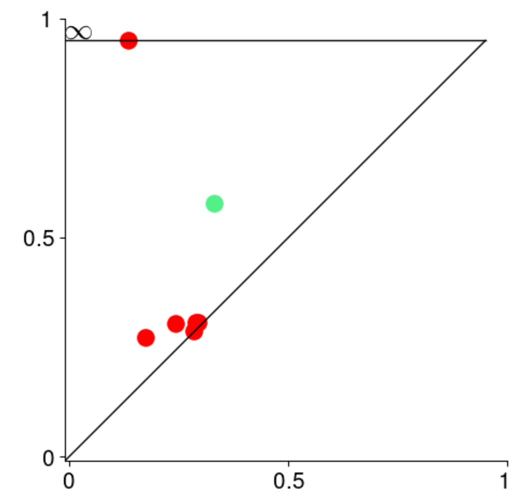
DTM-filtration, $p = 1$

$W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$



DTM-filtration, $p = 2$

$W^t[\mu, m, p] = \cup \overline{B}(x, (t^2 - d_{\mu,m}(x)^2)^{\frac{1}{2}})$



DTM-filtration, $p = 30$

$W^t[\mu, m, p] = \cup \overline{B}(x, (t^{30} - d_{\mu,m}(x)^{30})^{\frac{1}{30}})$

A stability result

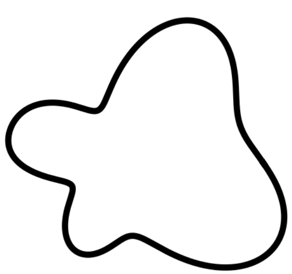
Case $p = 1$: $W^t[\mu, m, p] = \cup \overline{B}(x, t - d_{\mu,m}(x))$

Define $c(\mu, m, p) = \sup_{x \in \mathrm{supp}(\mu)} d_{\mu,m}(x)$.
Note: c is small if the measure μ is close to the Hausdorff measure restricted to a submanifold.

Theorem: Let μ, ν be probability measures. Let ν' be any probability measure with compact support included in $\mathrm{supp}(\nu)$. The interleaving distance between the (set) filtrations $W[\mu, m, p]$ and $W[\nu, m, p]$ is bounded by:

$$m^{\frac{1}{2}} W_2(\mu, \nu') + m^{\frac{1}{2}} W_2(\nu', \nu) + c(\mu, m, p) + c(\nu', m, p)$$

μ



ν'

ν

A stability result

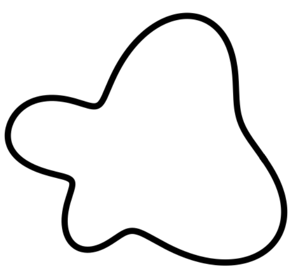
Case $p > 1$: $W^t[\mu, m, p] = \cup \overline{B}(x, (t^p - d_{\mu,m}(x)^p)^{\frac{1}{p}})$

Define $c(\mu, m, p) = \sup_{x \in \mathrm{supp}(\mu)} d_{\mu,m}(x) + 2(1 - \frac{1}{p})\mathrm{diam}(\mathrm{supp}(\mu))$.

Note: c is small if the measure μ is close to the Hausdorff measure rectricted to a submanifold.

Theorem: Let μ, ν be probability measures. Let ν' be any probability measure with compact support included in $\mathrm{supp}(\nu)$. The interleaving distance between ~~the (set) filtrations~~ the persistence modules $W[\mu, m, p]$ and $W[\nu, m, p]$ is bounded by:

$$m^{\frac{1}{2}} W_2(\mu, \nu') + m^{\frac{1}{2}} W_2(\nu', \nu) + c(\mu, m, p) + c(\nu', m, p)$$

μ 

ν' 

ν 

Conclusion

The DTM-filtrations $W[\mu, m, p]$ are Wassertein stable close to a submanifold.

Jupyter notebook available on my webpage
http://pages.saclay.inria.fr/raphael.tinarrage/

Thank you!