The DTM-filtration: a weighted filtration for persistent homology with noisy data and anomalous observations

Frédéric CHAZAL*, Marc GLISSE*, Raphaël TINARRAGE*, Yuhei UMEDA†, Hiroya INAKOSHI† and Hirokazu ANAI†
* Datashape, INRIA Saclay † Fujitsu Laboratories, AI Lab

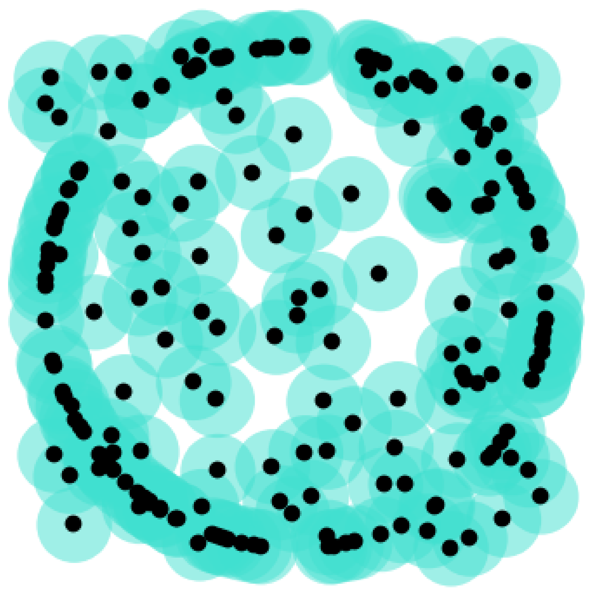
We present a new family of filtrations, the DTM-filtrations, which are a robust alternative to the filtrations commonly used in Topological Data Analysis. These DTM-filtrations allow us to reliably apply the methods of persistent homology, even with the presence of outliers. We apply this method to the detection of bridges cracks from acceleration data.

Definition and properties

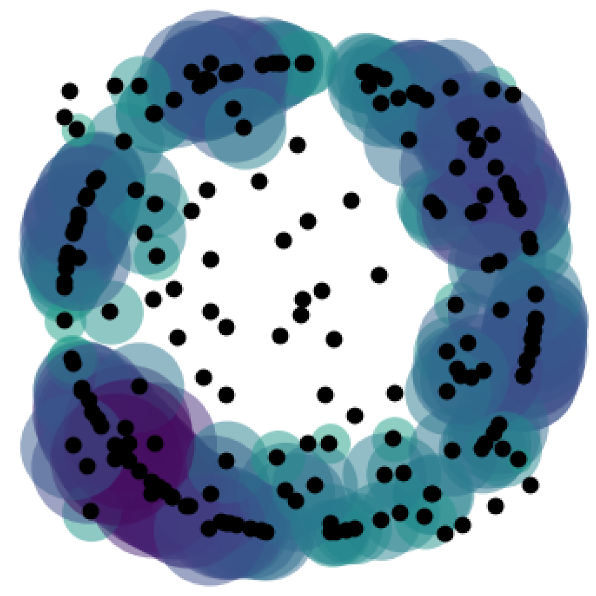
In the usual setting of persistent homology, one has a point cloud $X \subset \mathbb{R}^n$, and computes the persistence diagram $\mathrm{Diag}(\check{\mathrm{C}}\mathrm{ech}(X))$ of the Čech filtration of X. When X is close enough to a subset $K \subset \mathbb{R}^d$ in Hausdorff distance, one can then approximate the persistent homology of K :

$$d_B(\mathrm{Diag}(\check{\mathrm{C}}\mathrm{ech}(X)), \mathrm{Diag}(\check{\mathrm{C}}\mathrm{ech}(K)) \leq d_H(X, K).$$

However, as the Čech complex is sensitive to outliers, there is no chance to recover $\check{\mathrm{C}}\mathrm{ech}(K)$ from $\check{\mathrm{C}}\mathrm{ech}(X)$ if X contains anomalous points. We adress this issue with a Čech-like construction which takes into account the density of points.



The usual Čech filtration



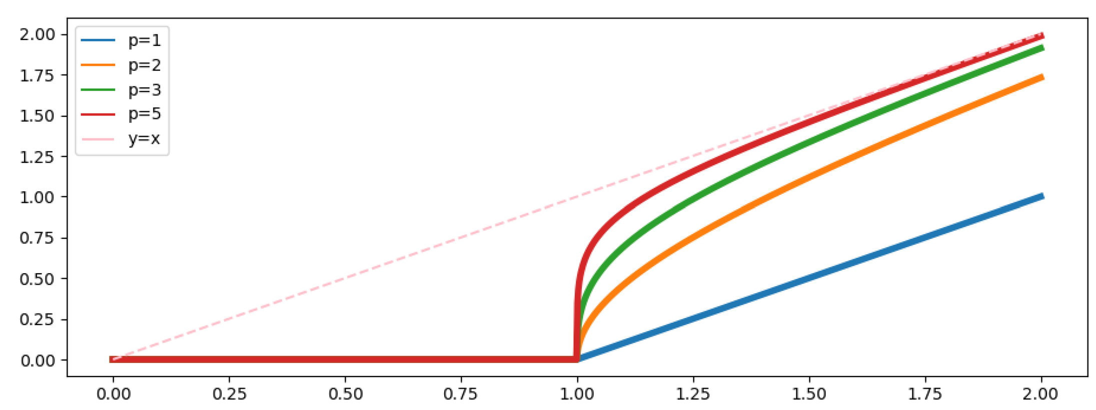
A DTM-filtration

Consider a continuous function $f : E \to \mathbb{R}^+$, and a real number $p \geq 1$. For every $x \in E$ and $t \in \mathbb{R}$, we define

$$r_x(t) = \begin{cases} -\infty & \text{if } t < f(x) \\ \left(t^p - f(x)^p\right)^{\frac{1}{p}} & \text{else.} \end{cases}$$

We denote by $\overline{B}_f(x,t) = \overline{B}(x, r_x(t))$ the closed euclidean ball of center x and radius $r_x(t)$. By convention, a ball of radius $-\infty$ is the empty set.

$t \mapsto r_x(t)$ for different values of p, and $f(x) = 1$.



Definition : weighted Čech filtration

Let $X \subset E$. We define the following set : $V^t = \bigcup_{x \in X} \overline{B}_f(x,t) \subset E$. The family $V = (V_t)_{t \geq 0}$ is a filtration of E. V is called the *weighted Čech filtration with parameters* (p, f, X), and is written $V = V_{p,f,X}$.

Proposition : Stability of weighted Čech filtrations

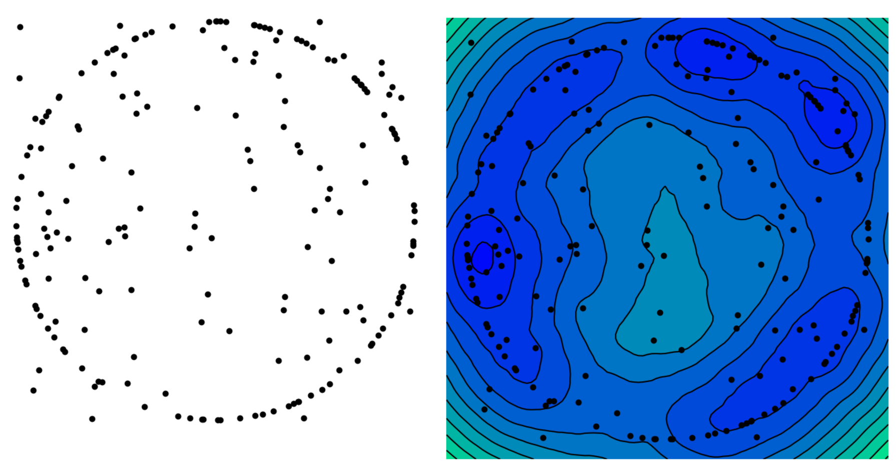
Consider two subsets $X, Y \subset E$, two functions $f, g : E \to \mathbb{R}^+$, and $p, q \geq 1$.

- $V_{p,f,X}$ and $V_{p,g,X}$ are ϵ-interleaved with $\epsilon = ||f - g||_\infty$.
- $V_{p,f,X}$ and $V_{p,f,Y}$ are $k\epsilon$-interleaved with $\epsilon = d_H(X,Y)$, $k = (1+c^p)^{\frac{1}{p}}$ and c a Lipschitz constant of f.
- $V_{p,f,X}$ and $V_{q,f,X}$ are ϵ-interleaved with $\epsilon = \sup_X(f)$.

Choice of f : the DTM

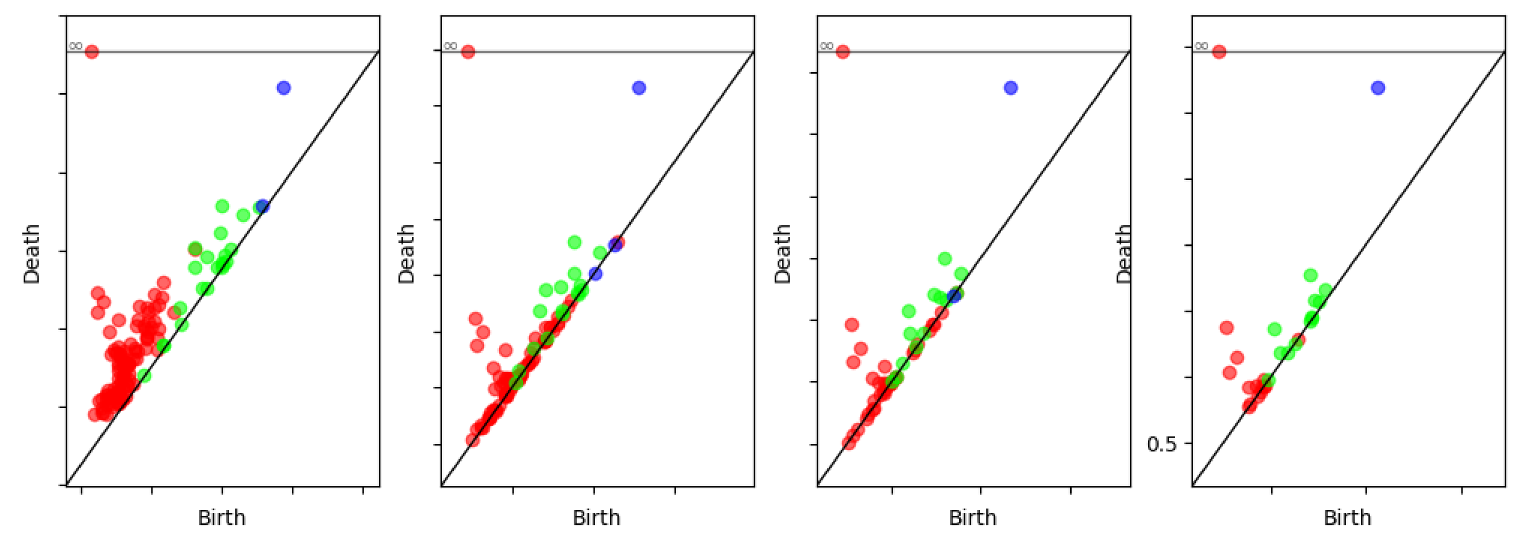
- Introduced in [1]
- Takes as inputs a probability measure over E and a real number $m \in [0,1)$, and returns a function $d_{\mu,m} : E \to \mathbb{R}^+$
- If $m = 0$, $d_{\mu,m} = d_{\mathrm{supp}(\mu)}$. If $m > 0$, it is a smoothing of it.

Contour plot of the DTM over a point cloud ($m = 0.1$)



Choice of p : sparsity

With X and f being fixed, the number of persistent H_0-cycles in $V_{p,X,f}$ is a non-increasing function of p. We also observe this behaviour for higher dimensional homology.



Persistence diagrams of $V_{p,X,f}$ for X sampled of the sphere and $p = 1, 2, 3, 5$.

Definition : the DTM-filtrations

Let $X \subset E$ be a finite point cloud, μ the empirical measure of X, $m \in [0,1)$ and $f = d_{\mu,m}$ the corresponding DTM. Let $p \geq 1$. Let V be the weighted Čech filtrations with parameters (p, f, X). We call it the *DTM-filtration associated with the parameters* (p, m, X), and denote it $V^{\mathrm{DTM}}_{p,m,X}$.

Proposition : Stability of the DTM-filtrations

Let X and Y be two finite subsets of E. Let $p \geq 1$ and $m \in [0,1)$. Consider the DTM-filtrations $V^{\mathrm{DTM}}_{p,m,X}$, $V^{\mathrm{DTM}}_{p,m,Y}$. We have

$$d_i(V^{\mathrm{DTM}}_{p,m,X}, V^{\mathrm{DTM}}_{p,m,Y}) \leq m^{-\frac{1}{2}} W_2(\mu_X, \mu_Y) + 2^{\frac{1}{p}} d_H(X,Y)$$

Proposition : Case $p = 1$

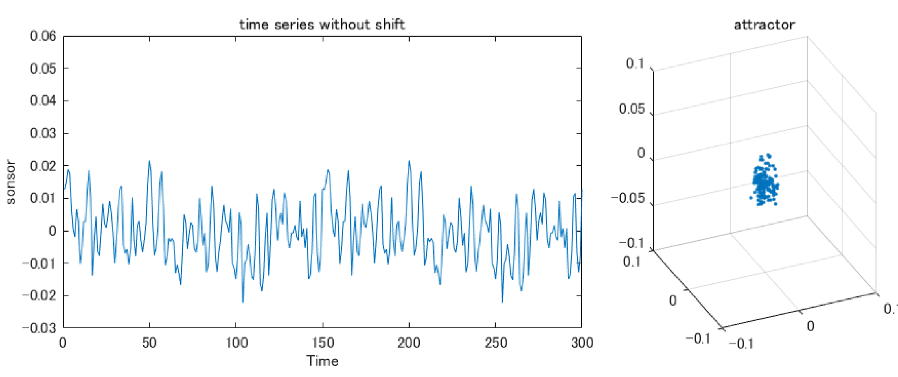
Let μ be a probability measure on E with compact support K. Let $m \in [0,1)$ and denote by V the sublevel sets filtration of $d_{\mu,m}$. Let $X \subset E$ finite, and define $V^{\mathrm{DTM}}_{1,m,X}$ the DTM-filtration and μ_X the empirical measure of X. Then

$$d_i(V, V^{\mathrm{DTM}}_{1,m,X}) \leq c + 2^{\frac{1}{p}}\epsilon + m^{-\frac{1}{2}} W_2(\mu, \mu_X)$$

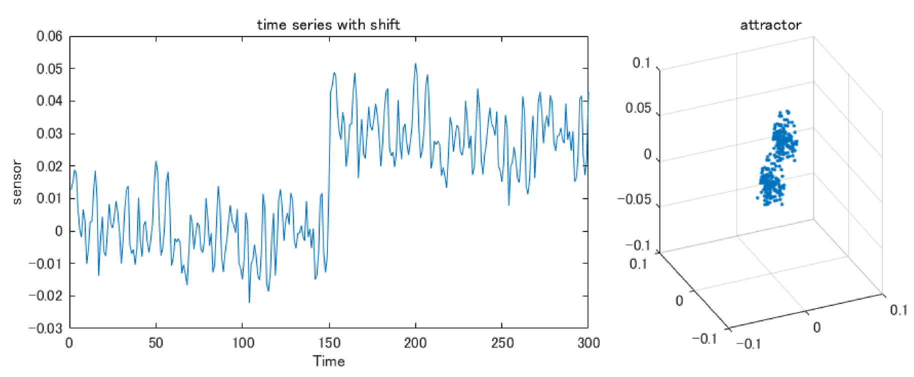
with $c = \sup_K(d_{\mu,m})$ and $\epsilon = d_H(K \cup X, X)$.

Application

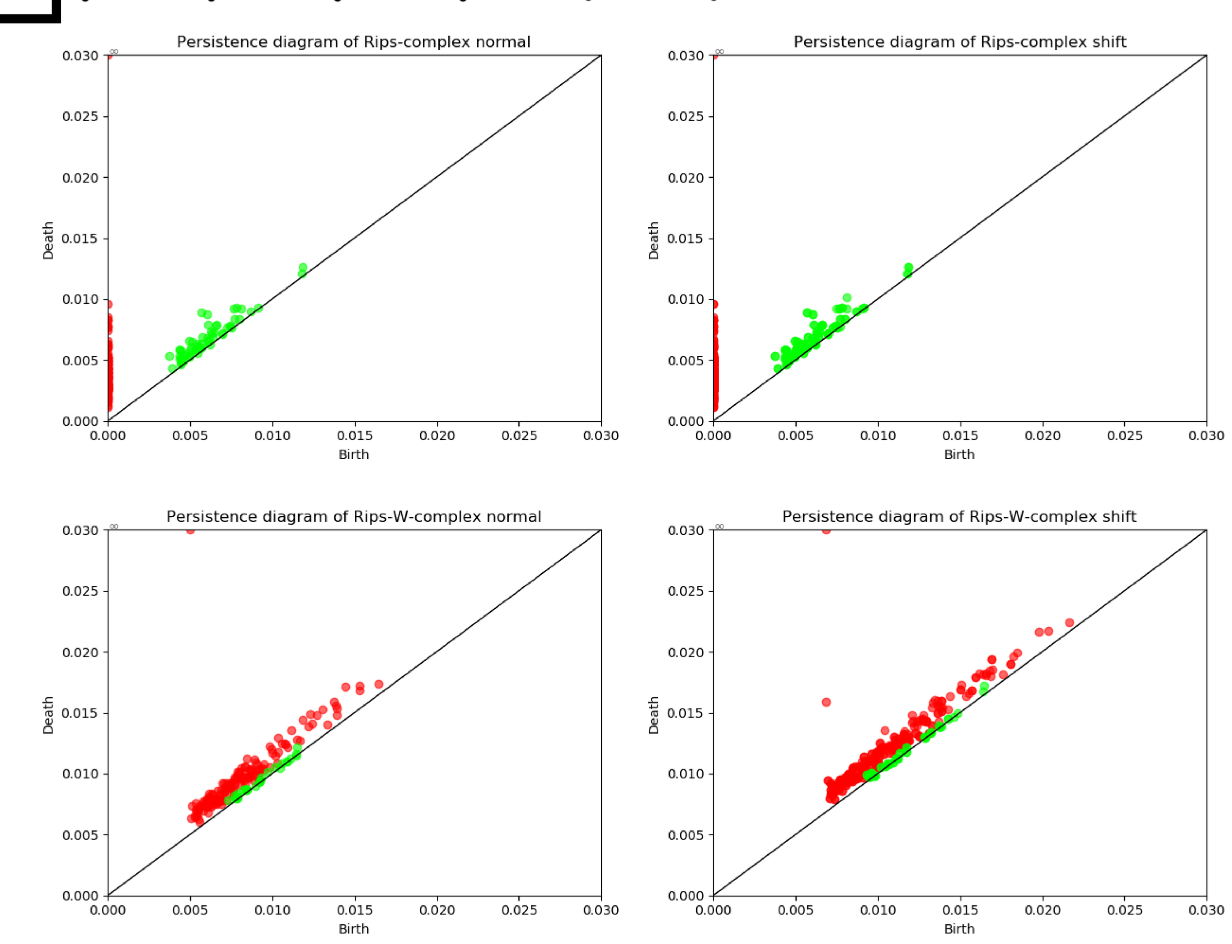
Crack-detection in bridges from acceleration sensors. We use a time-delay embedding.



Healthy area of the bridge



Internally damaged area



The usual Rips filtrations (persistence diagrams on the top) does not discriminate sufficiently between these two point clouds, while the DTM-filtration does (diagrams at the bottom).

[1] F. Chazal et al, Geometric Inference for Probability Measures, Fond. of Comp. Math.