Summary: Outlier Robustness

February 2, 2023

Abstract

This note is an attempt to summarise some of the articles and constructions we have been looking at in previous meetings. I will also try to summarise what kind of outlier robustness (leave-one-out stability) these methods are guaranteed to have.

Some fixed notation: Given a (finite) point cloud X with $|X| = n$, let μ_X denote the empirical probability measure with support on X. That is, $\mu_X = \frac{1}{n}\sum_{x\in X}\delta_x$ where δ_x denotes the Dirac measure centred at x. Similarly, let μ_{X_0} denote the empirical measure on $X_0 = X \setminus \{x_0\}$ for some arbitrary $x_0 \in X$.

1 Hausdorff distance and outliers

Stability in topological data analysis often refers to the robustness of topological features with respect to small perturbations in the data. We will refer to this type of stability as *classical stability*. However, classical stability is not sufficient to provide robustness against outliers. Given two subsets A and B of a metric space (Y, d), the *Hausdorff distance between A and B*, denoted $d_H(A, B)$ is defined as

$$d_H(A, B) = \inf\{\epsilon \geq 0 \mid A \subseteq B^\epsilon \text{ and } B \subseteq A^\epsilon\}$$

where $X^\epsilon = \{y \in Y \mid d(y, x) \leq \epsilon \text{ for some } x \in X\}$ is the ϵ-thickening of X with respect to d. Figure 1 illustrates the thickening of a point cloud sampled uniformly from the 2-torus embedded in $\mathbb{R}^3$.

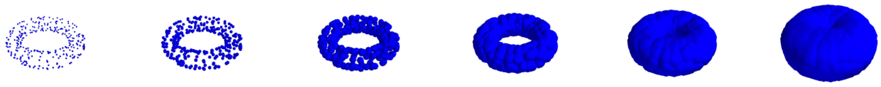


Figure 1: The ϵ-thickening of a point cloud sampled from a 2-dimensional torus. Here ϵ increases as we go from left to right.

For a point $y \in Y$ and a subset $X \subseteq Y$ we write $d(y, X) = \inf_{x \in X} d(y, x)$ and $d(-, X)$ for the function $y \mapsto d(y, X)$. The Hausdorff distance can then be written in the following two forms equivalent to the above definition:

$$d_H(A, B) = \max\left\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\right\} = \|d(-, A) - d(-, B)\|_\infty.$$

Given a function $f\colon Y \to \mathbb{R}$, we denote by $\mathrm{Dgm}(f)$ the persistence diagram of the sublevel sets of f. The persistence diagram $\mathrm{Dgm}(f)$ encodes the birth and death times of generators of $H_k(f^{-1}(-\infty, t])$ for all values of $t \in \mathbb{R}$ and integers $k \geq 0$. In other words, $\mathrm{Dgm}(f)$ records the topological features of the sublevels sets of f that persists when we vary t. There is a correspondence between persistence diagrams and what we call barcodes. There is a subtle difference in that barcodes contains information about whether or not the persistence intervals includes its endpoints. This difference can be important but in our case where we consider the homology of sublevel sets it does not matter so we view persistence diagrams and barcodes as representations of the same thing. Moreover, by the decomposition theorem in [5], every persistence module (of finite type) is completely characterized (up to isomorphism) by its corresponding barcode. The following result makes stability with respect to perturbation of the input data (or functions) precise:

Proposition 1.0.1 (Classical stability. See [2] and [4])**.** Let $f, g\colon Y \to \mathbb{R}$ be continuous and tame[1] functions with X triangulable. Then $d_I(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \|f - g\|_\infty$.

Now, for a point cloud $X \subseteq Y$, consider the distance function $f = d(-, X)$ and

[1]A function $f\colon Y \to \mathbb{R}$ is said to be *tame* if the homology groups of its sublevel sets are finite in all degrees and f has a finite number of homological critical values. See [2] for precise definitions.

observe that for any $t \geq 0$, we have

$$X^t := f^{-1}(-\infty, t] = \{y \in Y \mid \exists x \in X \text{ s.t. } d(x,y) \leq t\} = \bigcup_{x \in X} B(x,t).$$

In other words, the sublevels set of f are the thickenings of X. Since $X^t \subseteq X^s$ for $t \leq s$, we obtain a filtered topological space $\left\{X^t\right\}_{t \geq 0}$ called the *offset filtration* of X. Given a cover $\mathcal{U} = \{U_i\}_{i \in I}$ of a topological space Z, we define the *nerve of* $\mathcal{U}$, denoted $\mathcal{N}(\mathcal{U})$, to be the simplicial complex with vertex set I defined by

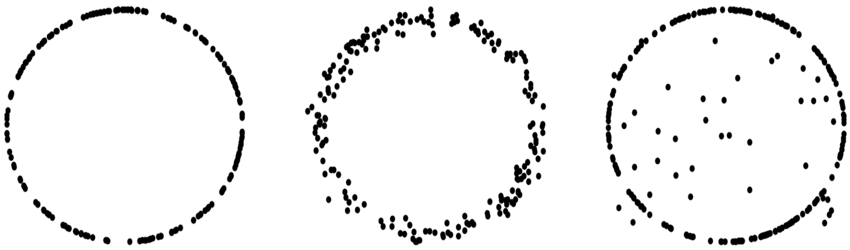
$$\sigma = \{i_0, i_1, \ldots, i_k\} \in \mathcal{N}(\mathcal{U}) \iff \bigcap_{i \in \sigma} \neq \emptyset$$

If for all finite non-empty $J \subseteq I$, the intersection $\bigcap_{i \in J} U_i$ is non-empty, the *nerve theorem* states that $\mathcal{N}(\mathcal{U})$ is homotopy equivalent to $Z = \bigcup_{i \in I} U_i$. The family $\{B(x,t)\}_{x \in X}$ defines a cover of X^t for every $t \geq 0$. Taking the nerve of the cover $\{B(x,t)\}_{x \in X}$, we obtain a simplicial complex $\check{\text{C}}\text{ech}_t(X)$ with vertex set X. Applying this construction for different values of t, we get a filtered simplicial complex $\check{\text{C}}\text{ech}(X) = \left\{\check{\text{C}}\text{ech}_t(X)\right\}_{t \geq 0}$ called the *Čech filtration of* X. By the nerve theorem, we have that the offset filtration and the Čech complex of X are homotopy equivalent, and hence they have isomorphic persistent homology modules. If we let $\text{Dgm}^{\check{\text{C}}}(X)$ denote the persistence diagram corresponding to the persistent homology of $\check{\text{C}}\text{ech}(X)$, the classical stability theorem can now be written as

$$d_I(\text{Dgm}^{\check{\text{C}}}(X), \text{Dgm}^{\check{\text{C}}}(Y)) \leq \|d(-,X) - d(-,Y)\|_\infty = d_H(X,Y).$$

1.0.1 Perturbation- vs. outlier stability

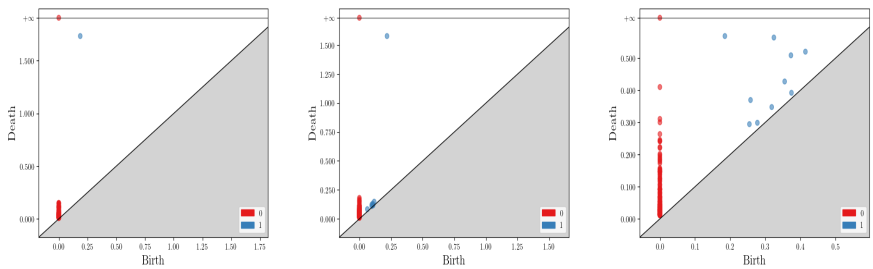
We now consider an simple example to demonstrate the above construction's sensitivity to outliers.



X_0 $\qquad$ X_1 $\qquad$ X_2

Figure 2: Three point clouds resembling a circle in the plane. The point cloud X_0 consists of points sampled uniformly from S^1, whereas X_1 is a slightly perturbed version of X_0, and X_2 is X_0 with a small number of outliers added.

The Hausdorff distance between X_0 and X_1 is small, whereas $d_H(X_0, X_2)$ is large in comparison. Figure 3 shows the corresponding persistence diagrams. As expected by the classical stability theorem, the diagrams corresponding to X_0 and X_1 are similar and both captures the homology of the underlying circle. However, the diagram of X_2 exhibits a lot of noise and it is not possible to infer the topology of a circle from the diagram.



$\mathrm{Dgm}^{\check{C}}(X_0)$ $\qquad$ $\mathrm{Dgm}^{\check{C}}(X_1)$ $\qquad$ $\mathrm{Dgm}^{\check{C}}(X_2)$

Figure 3: Three point clouds resembling a circle in the plane. The point cloud X_0 consists of points sampled uniformly from S^1, whereas X_1 is a slightly perturbed version of X_0, and X_2 is X_0 with a small number of outliers added.

1.0.2 The Vietoris-Rips complex

In general, the complexes in the Čech filtration is expensive to compute when the size of the input point cloud X is big and when our input data is high dimensional. In practice, one often use an approximation to the Čech filtration called the *Vietoris-Rips filtration*. Given a point cloud X, a k-simplex $\sigma = \{x_0, x_1, \dots, x_k\}$ is included in the Vietoris-Rips complex of X at filtration value $t \geq 0$, denoted $\text{Rips}_t(X)$, if and only if all pairwise distances $d(x_i, x_j) \leq t$. We can also define the Vietoris-Rips filtration as the sublevel set filtration of the diameter function $\sigma \mapsto \text{diam}(\sigma) = \max_{x_i, x_j \in \sigma}\{d(x_i, x_j)\}$. For point clouds in $\mathbb{R}^d$, it is shown in [3] that the Čech filtration and the Vietoris-Rips filtration are multiplicatively interleaved with interleaving constant $\sqrt{\frac{2d}{d+1}}$.

1.0.3 Previous work

What have been done to make PH robust against outliers? Specifically, density based methods. Constructions that takes the density in local neighbourhoods of X into consideration. Potential problems: (1) Computability in practice. (2) Additional parameters for which there is no obvious way to pick the best value. See Lesnick and Blumberg for a detailed review of previous methods ([10]).

2 Uniform stability

Let Y be a metric space, and consider a function V taking a finite subset $X \subseteq Y$ to a persistence module V_X. We introduce the notion of uniform stability for such functions with inspiration from learning theory (cf. [1]). If $X \subseteq Y$ is a point cloud and $x_0 \in X$, we fix the notation $X_0 = X \setminus \{x_0\}$. The intuition behind uniform stability is that, given enough samples, the difference between V_X and V_{X_0} should be very small.

Definition 2.1. We say that V is *uniformly stable with respect to d* if for every finite subset $X \subseteq Y$ with $|X| = n$, there exists a positive real number $\beta(n) \in O(n^{-1})$ such that

$$d\left(V_X, V_{X_0}\right) \leqslant \beta(n) \quad \text{holds for all } x_0 \in X. \tag{1}$$

Remark 2.1.1. The following three observations follows readily from the triangle inequality:

1. **Removal of multiple points:** Let $X_k = X \setminus \{x_0, x_1, \dots x_k\}$ for some $x_0, x_1, \dots, x_k \in X$. If V is uniformly stable, then $d\left(V_X, V_{X_k}\right)$ is also bounded by some $\beta(n) \in O(n^{-1})$.

2. **Replacing a point:** If $X' = (X \setminus \{x_0\}) \cup \{x'\}$ for some $x' \in Y$, then $d\left(V_X, V_{X'}\right)$ is also bounded by some $\beta(n) \in O(n^{-1})$.

3. **Interleavings and stability (?) :** If V and W are ϵ-interleaved and V is uniformly stable with respect to the interleaving distance d_I, then $d_I\left(W_X, W_{X_0}\right) \leq 2\epsilon + \beta(n)$.

Proof.

1. By the triangle inequality and uniform stability of V, we have that

$$d(V_X, V_{X_k}) \leq d(V_X, V_{X_{k-1}}) + d(V_{X_{k-1}}, V_{X_k}) \leq d(V_X, V_{X_{k-1}}) + \beta(n),$$

so the claim follows by induction.

2. Since $X_0 = X' \setminus \{x'\}$, it follows that

$$d(V_X, V_{X'}) \leq d(V_X, V_{X_0}) + d(V_{X' \setminus \{x'\}}, V_{X'}) \leq 2\beta(n).$$

3. From the assumptions we have $d_I(V_X, V_{X_0}) \leq \beta(n) \in O(n^{-1})$ and $d_I(V_Z, W_Z) \leq \epsilon$ for all Z. Using the triangle inequality, we have

$$\begin{aligned} d_I(W_X, W_{X_0}) &\leq d_I(W_X, V_X) + d_I(V_X, W_{X_0}) \\ &\leq \epsilon + d_I(V_X, V_{X_0})) + d_I(V_{X_0}, W_{X_0}) \leq 2\epsilon + \beta(n). \end{aligned}$$

□

2.1 The multicover bifiltration and the Prohorov distance

The *offset filtration* $O(X)$ of a point cloud X is a filtered topological space given in filtration degree $r > 0$ as the union of all balls of radius r centred at the points in X. Taking the nerve of this collection of balls of radius r, we get the usual Čech filtration. The multicover bifiltration $\mathcal{M}(X)$ is a 2-parameter extension of the offset filtration which is sensitive to density. Given a metric measure space (Y, η_Y), the *measure bifiltration* $\mathcal{B}(Y)$ is given in filtration degree (k, r) by

$$\mathcal{B}(Y)_{(k,r)} = \{y \in Y \mid \eta_Y(B(y, r)) \geq k\}.$$

For a point cloud $X \subseteq \mathbb{R}^d$, the *unormalized multicover bifiltration* $\mathcal{M}^u(X)$ of X is then defined as $\mathcal{M}^u(X) = \mathcal{B}(\hat{\nu}_X)$ where $\hat{\nu}_X$ is the counting measure on X defined as $\hat{\nu}_X(A) = |A \cap X|$. In other words, $\mathcal{M}^u(X)_{(k,r)}$ consists of all $y \in \mathbb{R}^d$ such that the ball centred at y of radius r contains at least k distinct points x from X. The *(normalized) multicover filtration* $\mathcal{M}(X)$ is defined similarly by replacing $\hat{\nu}_X$ with the empirical probability measure $\nu_X(A) = |A \cap X|/|X|$. This normalizes the first parameter k by the number of points in X.
The *Prohorov distance* d_{Pr} between two (probability) measures μ and η on a metric space (Y, d_Y) is given by

$$d_{Pr}(\mu, \eta) = \sup_A \inf\{\delta \geq 0 \mid \mu(A) \leq \eta(A^\delta) + \delta \text{ and } \eta(A) \leq \mu(A) + \delta\}$$

where A ranges over all closed sets in Y. The notation A^δ means the δ-thickening of A in Y with respect to the metric d_Y. It follows from remark 2.16 in [10] that the Prohorov distance between the two empirical probability measures μ_X and μ_{X_0} is bounded above by $\frac{1}{n-1}$. The following special case of a theorem of [10] then implies uniform stability for the multicover bifiltration with respect to the interleaving distance for 2-parameter persistence modules.

Theorem 2.2 (Theorem 1.6 [10])**.** For X and X' non-empty, finite subsets of $\mathbb{R}^d$,

$$d_I(\mathcal{M}(X), \mathcal{M}(X')) \leq d_{Pr}(\mu_X, \mu_{X'}).$$

Todo: Look at the subdivision and degree bifiltrations and understand how these are related to the multicover bifiltration. The paper [12] an example exploring stability of degree-Rips on an annulus with outliers.

Example 2.2.1. The Hausdorff distance between two point clouds is *not* bounded below by the Prohorov distance between their associated empirical measures. In other words, stability with respect to the Prohorov distance does not in general imply stability with respect to the Hausdorff distance. To see this, consider the following simple example: Given any $\epsilon \in (0, \frac{1}{2})$, let $X = \{x, y\}$ such that

$d(x, y) > 1$ and let X' be the union of X and a collection of points $\{x_1, x_2, \ldots, x_N\}$ with all $x_i \in B(y, \epsilon)$. Then, clearly the Hausdorff distance $d_H(X, X') \leq \epsilon$. Suppose that $\delta = d_{Pr}(\mu_X, \mu_{X'}) \leq \epsilon$. Then for $A = \{x\}$ we have $A^\delta = A$ so by the definition of the Prohorov distance the inequality $\mu_X(A) \leq \mu_{X'}(A^\delta) + \delta$ holds. This implies that $\frac{N+1}{N+2} \leq \delta \leq \epsilon$ which cannot be true since $\epsilon < \frac{1}{2}$. Hence, we must have that $d_{Pr}(\mu_X, \mu_{X'}) > d_H(X, X')$.

2.2 The Wasserstein distance

The Wasserstein distance, also known as the Earth Mover's Distance, is a distance measure between probability measures. It can be used to compare the similarity between two distributions, with smaller values indicating greater similarity. Some methods proposed in TDA for dealing with noise and outliers gives guarantees in terms of the Wasserstein distance. For two probability measures μ and η on $Y = \mathbb{R}^d$, we define the *p-Wasserstein distance* between μ and ν to be

$$W_p(\mu, \eta) = \inf_{\nu} \left(\int_{Y \times Y} \|x - y\|^p \, d\nu \right)^{\frac{1}{p}}.$$

Here, ν ranges over all couplings of μ and η. A coupling is a joint probability distribution on $Y \times Y$ with marginals equal to μ and η, respectively. Such couplings are also often referred to as transport plans, and one minimizing the above expression is called an optimal transport plan. For two measures μ and η with finite support, a transport plan can be represented as a matrix $A = (a_{ij})$ where a_{ij} is the mass moved from x_j to x_i. In this case, the p-Wasserstein distance between μ and η is given as

$$W_p(\mu, \eta) = \inf_{A} \left(\sum_{i,j} \|x_j - y_i\|^p \cdot a_{ij} \right)^{\frac{1}{p}}.$$

For example, let μ_X and μ_{X_0} be the empirical measures with support X and X_0, respectively. One way to think about transport plans between μ_X and μ_{X_0} is as follows: If we add a point x_0 to go from X_0 to X, each of the other $n-1$ points have to give away some mass so that all points, including x_0, ends up with a mass of $\frac{1}{n}$. One possible transport plan is given by the $(n-1) \times n$-matrix

$$\begin{bmatrix} \frac{1}{n(n-1)} & \frac{1}{n} & 0 & 0 & \dots & 0 & 0 \\ \frac{1}{n(n-1)} & 0 & \frac{1}{n} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n(n-1)} & 0 & 0 & 0 & \dots & \frac{1}{n} & 0 \\ \frac{1}{n(n-1)} & 0 & 0 & 0 & \dots & 0 & \frac{1}{n} \end{bmatrix}$$

Since the Wasserstein distance is the infimum over all such transport plans, we have the following upper bound

$$\begin{aligned} W_p(\mu_X, \mu_{X_0}) &\leq \left(\frac{1}{n(n-1)} \sum_{x \in X_0} \|x_0 - x\|^p \right)^{\frac{1}{p}} \\ &\leq \left(\frac{n-1}{n(n-1)} \operatorname{diam}(X)^p \right)^{\frac{1}{p}} = n^{-\frac{1}{p}} \operatorname{diam}(X). \end{aligned}$$

Idea:

Can we find a tighter bound for $d_W(\mu_X, \mu_{X_0})$?

Assuming that the diameter of X does not depend on $|X|$, the above inequality implies that the 1-Wasserstein distance between μ_X and μ_{X_0} is $O(n^{-1})$.

Section 5 in [9] says something about computing/approximating the Wasserstein distance for discrete measures.

2.2.1 The distance to a measure ([6])

For a probability measure μ on a metric space Y and a mass parameter $m \in (0, 1]$, define the function $\delta_{\mu,m} \colon \mathbb{R}^d \to \mathbb{R}_{>0}$ by sending a point y in $\mathbb{R}^d$ to the real number $\inf\{r > 0 \mid \mu(\bar{B}(y, r)) > m\}$. Intuitively, $\delta_{m,\mu}$ describes how large a disk around y we have to consider to capture at least m mass as defined by μ. Using δ, we define *the distance $d_{\mu,m}$ to measure μ* as

$$\begin{aligned} d_{\mu,m} \colon \mathbb{R}^d &\to \mathbb{R}_{>0} \\ y &\mapsto \sqrt{\frac{1}{m} \int_0^m \delta_{\mu,l}(y)^2 dl}. \end{aligned}$$

Let $\mathrm{Dgm}(d_{\mu,m})$ denote the persistence diagram corresponding to the sublevel set homology of $d_{\mu,m}$. In [6], the following Wasserstein stability result is established for the distance to a measure function:

Theorem 2.3 (Theorem 3.1 [6])**.** For two probability measures μ and ν and a triangulable metric space Y and a mass parameter m, we have that

$$d_B(\mathrm{Dgm}(d_{\mu,m}), \mathrm{Dgm}(d_{\nu,m})) \leq \frac{1}{\sqrt{m}} W_2(\mu, \nu).$$

In particular, for the empirical measures μ_X and μ_{X_0} this implies

$$d_B(\mathrm{Dgm}(d_{\mu_X,m}), \mathrm{Dgm}(d_{\mu_{X_0},m})) \leq \frac{1}{\sqrt{n}} \frac{1}{\sqrt{m}} \,\mathrm{diam}(X).$$

Approximation via a weighted Čech (or Rips) complex

Todo: Write something about stability in case of the complex used to approximate DTM. See [8]. There are also some stability results wrt. the Wasserstein distance in [8].

Computing the sublevel set homology of $d_{\mu,m}$ in practice is difficult. In [11] the authors propose a way to approximate the sublevel set homology by using weighted complexes.

2.2.2 Christoffel polynomials [11]

In the paper [11], the authors investigate use of Christoffel polynomials to achieve robustness to outliers in the input data. Given a point cloud X, one is interested in the sublevel set persistent homology module $\mathbb{CD}(X, k)$ of the logarithm[2] of the Christoffel polynomial $P_k^{\mu_X}$ of degree k induced by the empirical measure on X. Stability results are given with respect to the 1-Wasserstein measure. The following stability result is established in [11]:

Proposition 2.3.1. Let X and Y be point clouds in $[-1, 1]^d$. Then,

$$d_B\left(\mathrm{Dgm}(\mathbb{CD}(X, k)), \mathrm{Dgm}(\mathbb{CD}(Y, k))\right) \leq \log(C_X d_W(\mu_X, \mu_Y) + 1)$$

whenever $C_X \cdot d_W(\mu_X, \mu_Y) \leq 1$. Here $C_X = 2C_{d,k} \|P_k^{\mu_X}\|_\infty$ where $C_{d,k} = 4\binom{d+k}{k} k^2$.

[2]The sublevel sets of $\log P_k^{\mu_X}$ rather than $P_k^{\mu_X}$ are considered. The rationale given in the original paper is that this allows for stronger and more elegant stability results, and also that the logarithmic scaling produces persistence diagrams that better fit the underlying topology. See page 8 of [11].

Consider X and X_0 with their empirical measures μ_X and μ_{X_0}, respectively. Since $\operatorname{diam}[-1,1]^d = 2\sqrt{d}$ we have the upper bound $\log\left(2n^{-1}\sqrt{d}C_X + 1\right)$ for the bottleneck distance.

2.2.3 Kernel density estimation

Kernel density estimation is a non-parametric statistical technique for estimating the underlying probability density. It involves placing a kernel function, which is typically a smooth, continuous curve such as a normal distribution, at each data point and summing all of these kernel functions to estimate the true density function. The smoothness of the resulting curve is controlled by a parameter called the bandwidth. Sublevel set persistent homology can be applied to a kernel density estimator (KDE) to identify topological features in the data.

Definition 2.4. Let x be a point in Y. A *kernel g centred in x* is a function $g\colon Y \to \mathbb{R}_{>0}$ with $g(y) = h(\|x-y\|)$ for some continuous function $h\colon [0,\infty) \to \mathbb{R}_{>0}$. In addition, we require h to be differentiable on $(0,\infty)$ with $h'(t) < 0$. We call h the *height function* associated to g.

Given a kernel g centred in 0, we denote by g_x the shifted kernel given by $g_x(y) = g(x-y)$. If X is a finite point cloud, we define the *kernel density estimator (KDE) on X* (with kernel g) to be the function $f_X = \frac{1}{n}\sum_{x\in X} g_x\colon Y \to \mathbb{R}_{>0}$. Showing that sublevel set homology of a kernel density estimator is uniformly stable is straight-forward.

Proposition 2.4.1. The function taking a point cloud X to the sublevel set persistent homology of f_X is uniformly stable.

Proof. Let f_X and f_{X_0} be the kernel density estimators with kernel g on X and X_0, respectively.

$$f_X - f_{X_0} = \frac{1}{n} g_{x_0} - \frac{1}{n(n-1)} \sum_{x\in X_0} g_x = \frac{1}{n}\left(g_{x_0} - f_{X_0}\right).$$

Consequently, we have that $\|f_X - f_{X_0}\|_\infty \in O(n^{-1})$ and uniform stability follows from classical stability of persistence diagrams (cf. [2]). $\square$

A grid based approach to computing the sublevel set homology

One practical challenge with a KDE based approach is to compute the persistence homology. For low dimensions such as $d = 1$ and 2, a piecewise linear approximation using a grid can be used. However, this approach depends on the grid resolution, and the number of vertices increases exponentially in the number of dimensions which makes it unfeasible for higher dimensional data. The following approximation scheme is used by [11] to compute the sublevel set homology of a Lipschitz continuous function $f: [-1, 1]^d \to \mathbb{R}$:

1. Fix a positive integer m and construct the Freudenthal triangulation $\mathcal{K}_m$ of $[-1, 1]^d$.
2. This gives us a vertex set, with $(m+1)^d$ vertices, equal to the lattice points of $\frac{2}{m} \cdot \mathbb{Z}^n$.
3. Evaluate f on each vertex and compute the persistent homology module of the lower-star filtration on $\mathcal{K}_m$ induced by these function values. Denote the persistence diagram of this persistence module by $\mathrm{Dgm}(f|_{\mathcal{K}_m})$.

The following proposition gives an upper bound on the error of this approximation in terms of the Lipschitz constant for f:

Proposition 2.4.2 (Proposition 22 [11])**.** Let $f: [-1, 1]^d \to \mathbb{R}$ be Lipschitz continuous with Lipschitz constant L_f. Then

$$d_B\left(\mathrm{Dgm}(f|_{\mathcal{K}_m}), \mathrm{Dgm}(f)\right) \leq \frac{2L_f\sqrt{d}}{m}$$

Idea:

We know that all interesting topology of the sublevel sets of f_X is contained in the convex hull of X. If we consider a triangulation of the convex hull (e.g., the Delaunay triangulation or some refinement of it), can we give guarantees for the approximation using the lower-star filtration on this triangulation? (See proof and discussion in Appendix B of [11]).

Idea:

Can we first use a grid, and then apply some sparsification method (with approximation guarantees) to reduce the number of vertices?

2.2.4 The relationship between the multicover bifiltration and kernel density estimates

The multicover bifiltration is given in degree (r, k) as

$$\mathrm{Cov}_{r,k}(X) = \left\{ y \in \mathbb{R}^d \mid \|y - x\| \leq r \text{ for at least } k \text{ distinct } x \in X \right\}.$$

Now, consider the uniform kernel (also called the disk kernel or the boxcar kernel) $K\colon \mathbb{R} \to \mathbb{R}$ defined as $K(y) = \frac{1}{2}$ if $|y| \leq 1$ and $K(y) = 0$ otherwise. Given a bandwidth parameter r, we define the kernel density estimator with kernel K to be function $f_{X,r}\colon \mathbb{R}^d \to \mathbb{R}$ mapping $y \mapsto \frac{1}{n}\sum_{x \in X} K\left(\frac{\|x-y\|}{r}\right)$. The image of $f_{X,r}$ is precisely the set $\left\{\frac{k}{n}\right\}_{0 \leq k \leq n} \subset [0, 1]$. It is straightforward to see that the sublevel set $f_{X,r}^{-1}\left[\frac{k}{n}, \infty\right)$ of this density estimator is exactly the subspace $\mathrm{Cov}_{r,k}(X)$. Hence, we can view the multicover bifiltration as the superlevel sets of a KDE where we include the bandwidth as one of the persistence parameters.

2.2.5 Kernel distance to a measure

In [7], reconstruction properties and robustness of the kernel distance and kernel density estimates is shown. In this subsection, we follow the notation of [7]. A *kernel* is a function $K\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{>0}$ which takes high values on "similar" pair of points. For example, the Gaussian kernel defined by $K(p, x) = \sigma^2 \exp(-\frac{1}{2\sigma^2}\|p - x\|^2)$ (which is a scaled normal distribution with mean p). Given a point cloud $P \subset \mathbb{R}^d$, the *kernel density estimate* is defined as $\mathrm{KDE}_P(x) = \frac{1}{|P|}\sum_{p \in P} K(p, x)$. Now, given two point clouds P and Q, we define a similarity measure between them by letting $\kappa(P, Q) = \frac{1}{|P||Q|}\sum_{p \in P}\sum_{q \in Q} K(p, q)$ and define the *kernel distance* between P and Q (with respect to the kernel K) to be

$$D_K(P, Q) = \sqrt{\kappa(P, P) + \kappa(Q, Q) - 2\kappa(P, Q)}.$$

The similarity function κ can be generalized to a similarity function of measures. Given measures μ and ν, define $\kappa(\mu,\nu) = \int_{(p,q)\in\mathbb{R}^d\times\mathbb{R}^d} K dm_{\mu,\nu}(p,q)$ where $m_{\mu,\nu} = \mu\otimes\nu$ is the product measure of μ and ν. Consequently, we can define the kernel distance between measures by letting $D_K(\mu,\nu) = \sqrt{\kappa(\mu,\mu)+\kappa(\nu,\nu)-2\kappa(\mu,\nu)}$. If μ_P and μ_Q denote the empirical measures of P and Q, then $D_K(\mu_P,\mu_Q) = D_K(P,Q)$. The *kernel distance with respect to a measure* μ is defined as

$$d^K_\mu(y) = D_K(\mu,y) = \sqrt{\kappa(\mu,\mu)+\kappa(y,y)-2\kappa(\mu,y)}$$

and enjoys the following stability property:

$$\|d^K_\mu - d^K_\nu\|_\infty \le D_K(\mu,\nu).$$

For the empirical measures μ_X and μ_{X_0} we compute the square of the kernel distance to be

$$\begin{aligned}
D^2_K(\mu_X,\mu_{X_0}) &= \kappa(X,X)+\kappa(X_0,X_0)-2\kappa(X,X_0)\\
&= \frac{1}{n^2}\sum_{x\in X}\sum_{x'\in X}K(x,x') + \frac{1}{(n-1)^2}\sum_{x\in X_0}\sum_{x'\in X_0}K(x,x')\\
&\quad - \frac{2}{(n-1)n}\sum_{x\in X}\sum_{x'\in X_0}K(x,x')\\
&= \frac{1}{n^2}\left(K(x_0,x_0)+2\sum_{x\in X_0}K(x,x_0)+\sum_{x\in X_0}\sum_{x'\in X_0}K(x,x')\right)\\
&\quad + \frac{1}{(n-1)^2}\sum_{x\in X_0}\sum_{x'\in X_0}K(x,x')\\
&\quad - \frac{2}{(n-1)n}\left(\sum_{x\in X_0}\sum_{x'\in X_0}K(x,x')+\sum_{x\in X_0}K(x,x_0)\right)\\
&= \frac{1}{n^2}\left(K(x_0,x_0)+\frac{2}{n-1}\sum_{x\in X_0}K(x,x_0)+\frac{1}{(n-1)^2}\sum_{x\in X_0}\sum_{x'\in X_0}K(x,x')\right)
\end{aligned}$$

which shows that $D_K(\mu_X,\mu_{X_0}) \in O(n^{-1})$ and that the sublevel set homology of the kernel distance to the empirical measure μ_X is uniformly stable.

2.3 An overview of known stability results

Let X be a finte point cloud in $\mathbb{R}^d$ with $|X| = n$ and let $X_0 = X\setminus\{x_0\}$ for some arbitrary $x_0\in X$. Let μ_X and μ_{X_0} denote the empirical uniform measures on X and X_0, respectively.

2.3.1 Distance-to-measure (DTM)

Distance function $f: \mathbb{R}^d \to \mathbb{R}$
$f(y) = d_{\mu,m}(y) = \sqrt{\frac{1}{m}\int_0^m \delta^2_{\mu,t}(y)dt}$ where $\delta_{\mu,t}(y) = \inf\{r \geq 0 \mid \mu(\bar{B}(y,r)) \geq t\}$
Known upper bound [6]
$\|d_{\mu_X,m} - d_{\mu_{X_0},m}\|_\infty \leq \frac{1}{\sqrt{m}} W_2(\mu_X, \mu_{X_0}) \leq \frac{1}{\sqrt{nm}}\, \mathrm{diam}(X)$
Fixed parameter(s)
$m \in (0,1]$: mass parameter
Notes
Approximations of the sublevel persistent homology of $d_{\mu_X,m}$ have been proposed in [8] using weighted complexes.

2.3.2 Christoffel polynomials

Distance function $f: \mathbb{R}^d \to \mathbb{R}$
$f = \log P_k^{\mu}$ where P_k^{μ} is the Christoffel polynomial of degree k defined in terms of an orthonormal basis determined by the measure μ. See [11] for details.
Known upper bound [11]
$\|\log P_k^{\mu_X} - \log P_k^{\mu_{X_0}}\|_\infty \leq \log(C \cdot W_1(\mu_X, \mu_{X_0}) + 1) \leq \log\left(\frac{2C\sqrt{d}}{n} + 1\right)$
Fixed parameter(s)
$k \in \mathbb{N}$: polynomial degree
Notes
The constant C is defined as $C = C_{d,k} \max\left\{\|P_k^{\mu_X}\|_\infty, \|P_k^{\mu_{X_0}}\|_\infty\right\}$ with $C_{d,k} = 4\binom{d+k}{k}k^2$ where k is the polynomial degree and d is the ambient dimension. When it comes to computing the sublevel set persistent homology of the Christoffel polynomials, the authors of [11] construct a grid on $[-1,1]^d$ to linearly approximate $P_k^{\mu_X}$ and then compute the lower-star filtration.

2.3.3 Kernel density estimator (KDE)

Distance function $f\colon \mathbb{R}^d \to \mathbb{R}$
$f(y) = \frac{1}{n}\sum_{x\in X} K\left(\frac{\|x-y\|}{h}\right)$ where $K\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a kernel function.
Known upper bound
Fixed parameter(s)
$h > 0$: bandwidth
Notes

2.3.4 Kernel distance

Distance function $f\colon \mathbb{R}^d \to \mathbb{R}$
$f(y) = d_\mu^K(y) = D_K(\mu, y) = \sqrt{\kappa(\mu,\mu) + \kappa(y,y) - 2\kappa(\mu,y)}$.
Known upper bound
$\|d_{\mu_X}^K - d_{\mu_{X_0}}^K\|_\infty \leq D_K(\mu_X, \mu_{X_0}) \in O(n^{-1})$
Fixed parameter(s)
$\sigma > 0$: bandwidth
Notes
The sublevel sets of the kernel distance isomorphically maps to the superlevel sets of a kernel density estimate. Stability in terms of $D_K(X, X_0)$. It is shown that the kernel distance is distance-like and an approximation scheme involving weighted complexes (similar to the case of DTM) are used in the original paper.

References

[1] Olivier Bousquet and André Elisseeff. "Stability and generalization". In: *The Journal of Machine Learning Research* 2 (2002), pp. 499–526.

[2] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. "Stability of persistence diagrams". In: *Proceedings of the twenty-first annual symposium on Computational geometry*. 2005, pp. 263–271.

[3] Vin De Silva and Robert Ghrist. "Coverage in sensor networks via persistent homology". In: *Algebraic & Geometric Topology* 7.1 (2007), pp. 339–358.

[4] Frederic Chazal et al. *The structure and stability of persistence modules*. 2012. DOI: 10.48550/ARXIV.1207.3674. URL: https://arxiv.org/abs/1207.3674.

[5] William Crawley-Boevey. *Decomposition of pointwise finite-dimensional persistence modules*. 2012. DOI: 10.48550/ARXIV.1210.0819. URL: https://arxiv.org/abs/1210.0819.

[6] Mickael Buchet et al. *Efficient and Robust Persistent Homology for Measures*. 2013. DOI: 10.48550/ARXIV.1306.0039. URL: https://arxiv.org/abs/1306.0039.

[7] Jeff M. Phillips, Bei Wang, and Yan Zheng. *Geometric Inference on Kernel Density Estimates*. 2013. DOI: 10.48550/ARXIV.1307.7760. URL: https://arxiv.org/abs/1307.7760.

[8] Hirokazu Anai et al. *DTM-based Filtrations*. 2018. DOI: 10.48550/ARXIV.1811.04757. URL: https://arxiv.org/abs/1811.04757.

[9] Victor M. Panaretos and Yoav Zemel. "Statistical Aspects of Wasserstein Distances". In: *Annual Review of Statistics and Its Application* 6.1 (2019), pp. 405–431. DOI: 10.1146/annurev-statistics-030718-104938. URL: https://doi.org/10.1146%2Fannurev-statistics-030718-104938.

[10] Andrew J. Blumberg and Michael Lesnick. *Stability of 2-Parameter Persistent Homology*. 2020. DOI: 10.48550/ARXIV.2010.09628. URL: https://arxiv.org/abs/2010.09628.

[11] Pepijn Roos Hoefgeest and Lucas Slot. *The Christoffel-Darboux kernel for topological data analysis*. 2022. DOI: 10.48550/ARXIV.2211.15489. URL: https://arxiv.org/abs/2211.15489.

[12] Alexander Rolle. *The Degree-Rips Complexes of an Annulus with Outliers*. 2022. DOI: 10.48550/ARXIV.2203.08767. URL: https://arxiv.org/abs/2203.08767.