2. Machine Learning with Persistence Diagrams (MLwPD)

2.1. **Compact Geometric Filtrations.** Recall that a set $\mathcal{S} \subset \mathcal{D}$ is said to be:

(1) Bounded: If there exists a constant $C > 0$ so that $\mathsf{pers}(a, b) = b - a < C$ for every $(a, b) \in D$ and every $D \in \mathcal{S}$.
(2) Off Diagonally Birth Bounded (ODBB): If for every $\epsilon > 0$ there exists a constant $C_\epsilon > 0$ so that if $D \in \mathcal{S}$ and $(a, b) \in D$ satisfies $\mathsf{pers}(a, b) \geq \epsilon$, then we always have $b \leq C_\epsilon$.
(3) Uniformly Off Diagonally Finite (UODF): If for every $\epsilon > 0$ there exists a constant $M_\epsilon > 0$ so that for every $D \in \mathcal{S}$,

$$\#\left(\{\mathbf{x} \in D \ : \ \mathsf{pers}(\mathbf{x}) \geq \epsilon\}\right) \leq M_\epsilon.$$

Note that $\#(\cdot)$ above denotes cardinality of multisets (i.e., points are counted with multiplicity).

The characterization of compactness in $(\mathcal{D}, d_B)$ contends that $\mathcal{S} \subset \mathcal{D}$ is compact if and only if $\mathcal{S}$ is closed, bounded, ODBB and UODF. In the paper on machine learning with template functions we showed that if $\mathcal{S} \subset \mathcal{D}$ is compact and $F : \mathcal{S} \longrightarrow \mathbb{R}$ is continuous, then F can be approximated arbitrarily well using multivariate real polynomials and re-scaled/translets $f_j(\mathbf{x}) = f(a_j\mathbf{x}+\mathbf{b}_j)$, $a_j \in \mathbb{R}$ and $\mathbf{b}_j \in \mathbb{R}^2$, of any nonzero compactly-supported (template) function $f \in C(\mathbb{W}, \mathbb{R})$. The compactness of $\mathcal{S}$ is critical for this approach to work. The goal of this project is to identify which ways of generating filtered simplicial complexes from geometric data lead to sets of persistence diagrams guaranteed to form a compact subset of $\mathcal{D}$. That is, we want to know on what kinds of data sets it makes sense to apply templates.

A place to start: Let (Y, d_Y) be a finite metric space, and let $\mathsf{dgm}_j^{\mathcal{R}}(Y)$ denote the persistence diagram of the j-th persistent homology over a field $\mathbb{F}$, $j \geq 0$, of the Rips filtration on (Y, d_Y).

(1) Construct a sequence of finite subsets $Y_n \subset \mathbb{R}^\infty$ so that if Y_n is endowed with the ambient distance d_2 from $\mathbb{R}^\infty$, then $\mathsf{diam}(Y_n) \leq 2$ for all $n \in \mathbb{N}$, but $\left\{\mathsf{dgm}_1^{\mathcal{R}}(Y_n)\right\}_{n\in\mathbb{N}}$ is not compact. The goal here is to get a sense of how a family of data sets (the Y_n's) can fail to produce compact subsets of the space of persistence diagrams.

References.

(1) Approximating Continuous Functions on Persistence Diagrams Using Template Functions, J. A. Perea, L. Munch and F. Khasawneh https://arxiv.org/pdf/1902.07190.pdf
(2) Gromov's compactness theorem
(3) Proximity of Persistence Modules and their Persistence Diagrams, F. Chazal et. al., https://geometrica.saclay.inria.fr/team/Marc.Glisse/publications/futur/stable.pdf