THE PERSISTENT HOMOLOGY OF LIPSCHITZ EXTENSIONS

KIRK P. GARDNER AND DONALD R. SHEEHY

1. Introduction

Algorithms based on persistent homology form the core of many methods in topological data analysis. The standard approach is to compute a discrete representation of the persistent homology known as a *barcode* that captures the underlying topology. Theoretical guarantees on such algorithms are then given using the celebrated stability theory [3] which ensures that, for sufficiently good input data, the computed barcode will be close to the underlying ground truth. This is a strong guarantee on the output that necessarily requires strict conditions on the quality of the input data. If these conditions are not met, the algorithms provide no guarantees at all.

We present an application of a new approach to topological data analysis (TDA) using persistent homology with strong guarantees on the output even in the absence of strong sampling assumptions. Specifically, we will apply the recently developed theory of sub-barcodes [5] to Lipschitz extensions. Given a sample of an unknown Lipschitz function f, we give a simple algorithm for computing a sub-barcode of the barcode of f in general metric spaces.

> *The goal is to approximate a barcode that is a sub-barcode of all possible Lipschitz functions that agree with the input data.*

The primary algorithmic technique used is to compute the so-called image persistence [6, 2] associated with the sublevel sets of the maximum and minimum Lipschitz extensions of a sample. We will begin with the necessary background on metric spaces, homology, persistence modules, and barcodes in Section 2. In Section 3 we will review sub-barcodes, present the Lipschitz extension problem, and show how to approximate a sub-barcode in general metric spaces using the (Vietoris-)Rips complex.

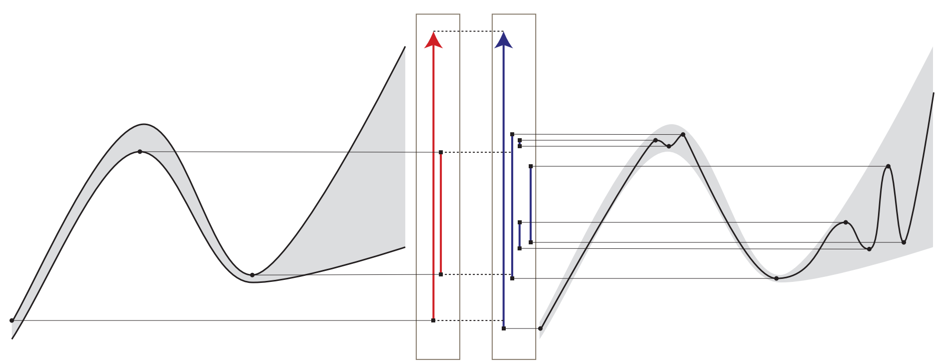

Figure 1. On the left, two functions are depicted: one is an upper bound and the other is a lower bound on an unknown function f, shown on the right. The barcode of the inclusion of the upper and lower bounds is a sub-barcode of $\mathcal{B}_f$.

2. Background

We will assume that the input is defined in a metric space (X, d). A function $f : X \to \mathbb{R}$ is c-Lipschitz if for all $x, y \in X$, we have $|f(x) - f(y)| \leq c\mathrm{d}(x, y)$. For $\varepsilon > 0$ let $\mathrm{ball}^{\varepsilon}(x) = \{y \in X \mid \mathrm{d}(x, y) < \varepsilon\}$ denote the metric ball centered at $x \in X$. A metric ball is said to be *strongly convex* if for each pair of points y, z in its closure, there exists a unique shortest path in X between y and z whose interior is contained in the metric ball. For $x \in X$, let $\varrho_X(x)$ be the supremum of the radii ε such that $\mathrm{ball}^\varepsilon(x)$ is strongly convex. The **strong convexity radius** of X is defined as $\varrho_X = \inf_{x\in X} \varrho_X(x)$.

The distance of a point $x \in X$ to a set $S \subseteq X$ is defined as $\mathrm{d}(x, S) = \inf_{s\in S} \mathrm{d}(x, s)$. The **Hausdorff distance** between two sets $S, T \subseteq X$ is defined as $\mathrm{d}_H(S, T) = \max\left\{ \sup_{s\in S} \mathrm{d}(s, T),\ \sup_{t\in T} \mathrm{d}(S, t)\right\}$.

2.1. **Homology.** We will assume the reader is familiar with homology, and refer to Hatcher [7] for a full treatment. Throughout, we will be using singular homology over a field k so that the homology groups $\mathsf{H}_n(X)$ of a topological space X are vector spaces for all $n \in \mathbb{N}$. We will write $\mathsf{H}(X)$ to refer to the homology of X for any dimension n. For any continuous map $f : X \to Y$, there is a corresponding linear map $\mathsf{H}[f] : \mathsf{H}(X) \to \mathsf{H}(Y)$. In particular, let $\mathsf{H}[X \subseteq Y]$ denote the map in homology induced the inclusion of a subspace $X \subseteq Y$.

2.2. **Filtrations.** A **filtration** is a nested family of topological spaces. In this work, a filtration will be defined as a mapping $F : \mathbb{R} \to \mathbf{Top}$ associating a topological space $F(t)$ to each $t \in \mathbb{R}$ such that $F(s) \subseteq F(t)$ for all $s \le t$. There is a natural partial order on filtrations where $F \hookrightarrow G$ if and only if $F(t) \subseteq G(t)$ for all $t \in \mathbb{R}$. Given a space X and a function $f : X \to \mathbb{R}$, the **sublevel filtration** of f is defined $\mathrm{Sub}(f) = \left\{f^{-1}((-\infty, t])\right\}_{t\in\mathbb{R}}$ where $f^{-1}((-\infty, t]) = \left\{x \in X \mid f(x) \le t\right\}$. The set of real-valued functions $X \to \mathbb{R}$ forms a partially ordered set where $g \ge f$ if $g(x) \ge f(x)$ for all $x \in X$. It follows from these definitions that $g \ge f$ implies that $\mathrm{Sub}(g) \hookrightarrow \mathrm{Sub}(f)$.

2.3. **Persistence Modules.** A **persistence module** $\mathbb{V}$ (over $\mathbb{R}$) consists of an $\mathbb{R}$-indexed family of vector spaces $\mathbb{V}(t)$ for $t \in \mathbb{R}$ and linear maps $\mathbb{V}[s \le t] : \mathbb{V}(s) \to \mathbb{V}(t)$ for $s \le t$ subject to the following conditions.

(1) $\mathbb{V}(t \le t) = 1_{\mathbb{V}(t)}$ for all t and
(2) $\mathbb{V}(s \le t) \circ \mathbb{V}(r \le s) = \mathbb{V}(r \le t)$ for all $r \le s \le t$.

A **homomorphism** $\phi : \mathbb{V} \to \mathbb{W}$ of persistence modules is an $\mathbb{R}$-indexed collection of linear maps $\phi = \left\{\phi_t : \mathbb{V}(t) \to \mathbb{W}(t)\right\}_{t\in\mathbb{R}}$ such that $\phi_t \circ \mathbb{V}[s \le t] = \mathbb{W}[s \le t] \circ \psi_s$ for all $s \le t$. A persistence module $\mathbb{V}$ is **pointwise finite-dimensional (p.f.d.)** if $\mathbb{V}(t)$ is finite-dimensional for all $t \in \mathbb{R}$.

Convention 2.1. For the purposes of this work, we will assume that all persistence modules are p.f.d.

The most common way to produce a persistence module $\mathbb{V}$ is to consider the homology of a filtration F. That is, $\mathsf{H}F$ is a persistence module consisting of vector spaces $\mathsf{H}(F(t))$ and linear maps $\mathsf{H}[F(s) \subseteq F(t)] : \mathsf{H}(F(s)) \to \mathsf{H}(F(t))$.

2.4. **Barcodes.** Let **Int** denote the set of intervals in $\mathbb{R}$. A **barcode** is a function $\mathcal{B} : \underline{\mathcal{B}} \to \mathbf{Int}$ that associates each *bar* β in the set $\underline{\mathcal{B}}$ with an interval $\mathcal{B}(\beta) \in \mathbf{Int}$. A persistence module may be constructed from a barcode as the direct sum of *interval modules* that is unique up to isomorphism. That is, the barcode of a p.f.d. persistence module $\mathbb{V}$ is unique up to isomorphism, and will be denoted $\mathcal{B}_\mathbb{V}$ or $\mathcal{B}(\mathbb{V})$ when no confusion will occur.

Notations 2.1. (i) For any homomorphism $\phi : \mathbb{V} \to \mathbb{W}$ of p.f.d. persistence modules, the image of ϕ is a p.f.d. persistence module, and therefore has a barcode that will be denoted $\mathcal{B}_\phi = \mathcal{B}_{\mathsf{im}\,\phi}$.
(ii) For any filtration F, let $\mathcal{B}_F = \mathcal{B}_{\mathsf{H}F}$ denote the barcode of the persistent homology module $\mathsf{H}F$.
(iii) For any function $f : X \to \mathbb{R}$, let $\mathcal{B}_f = \mathcal{B}(\mathsf{H}\mathrm{Sub}(f))$ denote the barcode of the persistent homology of the sublevel filtration of f, and let $\mathcal{B}(g \ge f) = \mathcal{B}(\mathsf{H}\mathrm{Sub}[g \ge f])$ denote the barcode of the image of the map in homology induced by the inclusion of filtrations $\mathrm{Sub}(g) \hookrightarrow \mathrm{Sub}(f)$.

Let $\delta \ge 0$ and for any interval $I \in \mathbf{Int}$ let $\mathsf{S}^\delta(I) = \left\{t \in I \mid t - \delta \in I \text{ and } t + \delta \in I\right\}$. Given a barcode $\mathcal{B} : \underline{\mathcal{B}} \to \mathbf{Int}$ let $\underline{\mathcal{B}}^\delta = \left\{\beta \in \underline{\mathcal{B}} \;\middle|\; \mathsf{S}^\delta(\mathcal{B}(\beta)) \ne \emptyset\right\}$. The **$\delta$-smoothing** of $\mathcal{B}$ is the barcode $\mathcal{B}^\delta : \underline{\mathcal{B}}^\delta \to \mathbf{Int}$ defined for $\beta \in \underline{\mathcal{B}}^\delta$ as $\mathcal{B}^\delta(\beta) = \mathsf{S}^\delta(\mathcal{B}(\beta))$. A **$\delta$-bottleneck matching** between barcodes $\mathcal{A}$ and $\mathcal{B}$ is a subset $M \subseteq \underline{\mathcal{A}} \times \underline{\mathcal{B}}$ such that

(1) for all $\alpha \in \underline{\mathcal{A}}^\delta$ there is a unique $\beta \in \underline{\mathcal{B}}$ with $(\alpha, \beta) \in M$ and $\mathcal{A}^\delta(\alpha) \subseteq \mathcal{B}(\beta)$;
(2) for all $\beta \in \underline{\mathcal{B}}^\delta$ there is a unique $\alpha \in \underline{\mathcal{A}}$ with $(\alpha, \beta) \in M$ and $\mathcal{B}^\delta(\beta) \subseteq \mathcal{A}(\alpha)$.

The **bottleneck distance** between $\mathcal{A}$ and $\mathcal{B}$ is defined

$$\mathrm{d}_B(\mathcal{A}, \mathcal{B}) = \inf\left\{\delta \ge 0 \mid \text{there exists a } \delta\text{-bottleneck matching between } \mathcal{A} \text{ and } \mathcal{B}\right\}.$$

3. Sub-Barcodes and The Lipschitz Extension Problem

In this section we review the theory of sub-barcodes from the work of Chubet et al. [5] which gives sufficient conditions in order to guarantee that the barcode of a persistence module is a sub-barcode of another. Here, we present a special case that will be used to prove the correctness of the algorithms in this paper.

Definition 3.1 (Sub-barcode)**.** Given barcodes $\mathcal{A}$ and $\mathcal{B}$, we say that $\mathcal{A}$ is a sub-barcode of $\mathcal{B}$ and write $\mathcal{A} \sqsubseteq \mathcal{B}$ if there is an injective map $M : \underline{\mathcal{A}} \to \underline{\mathcal{B}}$ such that for all $\alpha \in \underline{\mathcal{A}}$, we have $\mathcal{A}(\alpha) \subseteq \mathcal{B}(M(\alpha))$. Note that $\mathcal{B}^\delta \sqsubseteq \mathcal{B}$ and $\mathrm{d}_B(\mathcal{A}, \mathcal{B}) \leq \delta$ implies $\mathcal{A}^\delta \sqsubseteq \mathcal{B}$ and $\mathcal{B}^\delta \sqsubseteq \mathcal{A}$.

Cohen-Steiner et al. [6] showed that the barcode of a homomorphism induced by an inclusion of filtrations can be computed efficiently. More recently, Bauer and Schmahl [2] gave an efficient implementation incorporating several heuristics. These results provide a natural computational variant of the following theorem which will be used to establish guarantees on the algorithms presented in the following section. The proof is based on the induced matching theory of Bauer and Lesnick [1].

Theorem 3.2. *If $\phi : \mathbb{U} \to \mathbb{V}$ and $\psi : \mathbb{V} \to \mathbb{W}$ are persistence module homomorphisms then $\mathcal{B}_{\psi\phi} \sqsubseteq \mathcal{B}_{\mathbb{V}}$.*

The utility of Theorem 3.2 comes from the following corollary for ordered functions.

Corollary 3.3. *If $g \geq f \geq \ell : X \to \mathbb{R}$ then $\mathcal{B}(g \geq \ell) \sqsubseteq \mathcal{B}_f$.*

3.1. The Lipschitz Extension Problem.

Let X and Y be metric spaces and let $f : X \to Y$ be an unknown Lipschitz function. In the Lipschitz extension problem, the input is a finite sample P of X, in addition to function values $f(p)$ for each sample point $p \in P$; the output is a Lipschitz function $\tilde{f}_P : X \to Y$ such that $f(p) = \tilde{f}_P(p)$ for all $p \in P$. In the special case where the codomain of f is $\mathbb{R}$, it is always possible to define such an extension. In fact, there are two canonical Lipschitz extensions of real-valued functions that we will consider in this paper.

Definition 3.4 (Maximum and Minimum Lipschitz Extensions)**.** Let X be a metric space and let $f : X \to \mathbb{R}$ be an unknown c-Lipschitz function. Given a finite sample P of X let $f_P = f|_P : P \to \mathbb{R}$ denote the restriction of f to P. The **minimum Lipschitz extension** $f_P^\wedge : X \to \mathbb{R}$ of f_P is defined for $x \in X$ as

$$f_P^\wedge(x) = \max_{p \in P} f(p) - c\mathrm{d}(x, p).$$

Similarly, the **maximum Lipschitz extension** $f_P^\vee : X \to \mathbb{R}$ is defined

$$f_P^\vee(x) = \min_{p \in P} f(p) + c\mathrm{d}(x, p).$$

It follows directly from these definitions that $f_P^\vee \geq f \geq f_P^\wedge$ for any sample $P \subset X$. The notation $f_P^\vee$ and $f_P^\wedge$ is intended to evoke the notation for joins and meets in lattices.

3.2. A Vietoris-Rips Approach for Metric Data.

In order to do computations, a filtration is often modeled as a discrete object known as a simplicial complex. Formally, a **simplicial complex** K is a pair (V, S) where V is any set and S is a collection of subsets $\sigma \subseteq S$ known as *simplices* such that $\sigma \in S$ and $\tau \subseteq \sigma$ implies $\tau \in S$. A *filtered simplicial complex* is a filtration F in which $F(t)$ is a simplicial complex for all $t \in \mathbb{R}$.

In particular, we will make use of the **(Vietoris-)Rips complex**, which is defined for a finite sample $P \subset X$ and $\varepsilon > 0$ as the simplicial complex $\mathcal{R}^\varepsilon(P) = \left\{\sigma \subseteq P \,\middle|\, \mathrm{d}(p, q) \leq \varepsilon \text{ for all } p, q \in \sigma\right\}$. The Rips complex can be easily computed from metric data, however, it does not accurately capture the topology of the underlying space. On the other hand, the **Čech complex**, defined $\check{\mathcal{C}}^\varepsilon(P) = \left\{\sigma \subseteq P \,\middle|\, \bigcap_{p \in \sigma} \mathrm{ball}^\varepsilon(p) \neq \emptyset\right\}$, enjoys a homotopy equivalence with the corresponding metric cover, under suitable sampling conditions, by the Nerve Theorem. Formally, we will assume the strong convexity radius ϱ_X of X is sufficiently large so that S will form a good open cover. Importantly, we have the following sequence of inclusions referred to as the *Rips-Čech interleaving*:

$$\mathcal{R}^\varepsilon(P) \subseteq \check{\mathcal{C}}^\varepsilon(P) \subseteq \mathcal{R}^{2\varepsilon}(P).$$

Using this fact, we will approximate the topology of the Čech complex, and therefore the underlying metric cover, using a pair of Rips complexes $\mathcal{R}^\varepsilon(P) \hookrightarrow \mathcal{R}^{2\varepsilon}(P)$ as follows.

For any function $f : X \to \mathbb{R}$, let S_t denote the sublevels of f_S. For any $\varepsilon \geq 0$, let $\mathcal{R}_f^\varepsilon(S)$ and $\check{\mathcal{C}}_f^\varepsilon(S)$ respectively denote the filtrations on $\mathcal{R}^\varepsilon(S)$ and $\check{\mathcal{C}}^\varepsilon(S)$ induced by f_S defined

$$\mathcal{R}_f^\varepsilon(S) = \left\{\mathcal{R}^\varepsilon(S_t)\right\}_{t\in\mathbb{R}} \text{ and } \check{\mathcal{C}}_f^\varepsilon(S) = \left\{\check{\mathcal{C}}^\varepsilon(S_t)\right\}_{t\in\mathbb{R}}.$$

Theorem 3.5. *Let X be compact metric space with $\varrho_X > 2\delta$ and let $S \subseteq X$ with $\mathrm{d}_H(X, S) \leq \delta$. If $g \geq f \geq \ell : X \to \mathbb{R}$ then*

$$\mathcal{B}^{c\delta}\big(\mathcal{R}_g^\delta(S) \hookrightarrow \mathcal{R}_\ell^{2\delta}(S)\big) \sqsubseteq \mathcal{B}_f.$$

Proof. By the standard Rips-Čech interleaving and the function ordering, we have

$$\mathcal{R}_g^\delta(S) \hookrightarrow \check{\mathcal{C}}_g^\delta(S) \hookrightarrow \check{\mathcal{C}}_f^\delta(S) \hookrightarrow \check{\mathcal{C}}_\ell^\delta(S) \hookrightarrow \mathcal{R}_\ell^{2\delta}(S).$$

Thus, by Corollary 3.3, we get $\mathcal{B}\big(\mathcal{R}_g^\delta(S) \hookrightarrow \mathcal{R}_\ell^{2\delta}(S)\big) \sqsubseteq \mathcal{B}_{\check{\mathcal{C}}_f^\delta(S)}$.

Letting W_f denote the filtration defined for $t \in \mathbb{R}$ as $W_f(t) = \bigcup_{s\in S_t} \mathrm{ball}^\delta(s)$, we have that $\mathcal{B}_{W_f} = \mathcal{B}_{\check{\mathcal{C}}_f^\delta(S)}$ by the Persistent Nerve Lemma (see Chazal et al. [4], Lemma 3.4). Using the triangle inequality and the Lipschitz condition on f, we have $\mathrm{d}_B(\mathcal{B}_{W_f}, \mathcal{B}_f) \leq c\delta$, so $\mathcal{B}_{W_f}^{c\delta} \sqsubseteq \mathcal{B}_f$.

Putting these facts together implies the desired result that

$$\mathcal{B}^{c\delta}\big(\mathcal{R}_g^\delta(S) \hookrightarrow \mathcal{R}_\ell^{2\delta}(S)\big) \sqsubseteq \mathcal{B}_f.$$

□

The preceding theorem supports the following algorithm in the case where the function values are only known at a subset P of S. There is no assumption about the density of P.

The Vietoris-Rips Sub-barcode Algorithm:

Input: Constants δ and c, a δ-sample S of X, a subset P of S, and $f_P : P \to \mathbb{R}$.

Output: A barcode that is guaranteed to be a sub-barcode of every c-Lipschitz function on X that agrees with the input on P.

(1) Compute $\mathcal{R}^\delta(S)$ and $\mathcal{R}^{2\delta}(S)$.
(2) Compute the maximum Lipschitz extension $f_P^\vee$ on the vertices S.
(3) Similarly, compute the minimum Lipschitz extension $f_P^\wedge$ on the vertices.
(4) Compute the image persistence barcode $\mathcal{B}\big(\mathcal{R}_{f_P^\vee}^\delta(S) \hookrightarrow \mathcal{R}_{f_P^\wedge}^{2\delta}(S)\big)$.
(5) Return the $c\delta$-smoothing $\mathcal{B}^{c\delta}\big(\mathcal{R}_{f_P^\vee}^\delta(S) \hookrightarrow \mathcal{R}_{f_P^\wedge}^{2\delta}(S)\big)$.

Remark 3.1. The preceding constructions only require a sufficiently dense sample of X. It is possible to subsample S in a way that preserves this guarantee, but also reduces the complexity of the resulting complexes if the intrinsic dimension is low. Let d denote the doubling dimension of X, defined as the log of the maximum number of radius r metric balls needed to cover a ball of radius $2r$. Sheehy [8] showed that the complexity of a Vietoris-Rips complex at scale δ on an $O(\delta)$-net of X has total complexity $2^{O(d^2)}n$. This result was used to build an filtration to approximate the Vietoris-Rips complex at all scales. In the current work, it suffices to consider two different scales, and thus, two different linear size complexes. So, for low-dimensional input, there exists a linear size nested pair of filtrations whose image persistence gives a guaranteed sub-barcode.

References

[1] U. Bauer and M. Lesnick. Persistence diagrams as diagrams: A categorification of the stability theorem. In *Topological Data Analysis*, pages 67–96. Springer, 2020.
[2] U. Bauer and M. Schmahl. Efficient computation of image persistence, 2022.
[3] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics. Springer International Publishing, 2016.
[4] F. Chazal and S. Y. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proceedings of the 24th ACM Symposium on Computational Geometry*, pages 232–241, 2008.
[5] O. A. Chubet, K. P. Gardner, and D. R. Sheehy. A theory of sub-barcodes, 2022.
[6] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov. Persistent homology for kernels, images, and cokernels. In *SODA: ACM-SIAM Symposium on Discrete Algorithms*, 2009.
[7] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2001.
[8] D. R. Sheehy. Linear-size approximations to the Vietoris-Rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.