

Fréchet-Stable Signatures Using Persistence Homology*

Donald R. Sheehy
University of Connecticut
don.r.sheehy@gmail.com

Abstract

For a metric space Y , the Fréchet distance is a metric on trajectories $f, g : [0, 1] \rightarrow Y$ that minimizes $\max_{t \in [0, 1]} d_Y(f(t), g(h(t)))$ over continuous reparameterizations h of time. One can define the generalized Fréchet distance between more complex objects, functions $f : X \rightarrow Y$ where X is some topological space that minimizes over homeomorphisms from $X \rightarrow X$. This more general definition has been studied for surfaces and often leads to computationally hard problems. We show how to compute in polynomial-time signatures for these functions for which the resulting metric on the signatures can also be computed in polynomial-time and provides a meaningful lower bound on the generalized Fréchet distance. Our approach uses persistent homology and exploits the natural invariance of persistence diagrams of functions to homeomorphisms of the domain. Our algorithm for computing the signatures in Euclidean spaces uses a new method for computing persistent homology of convex functions on simplicial complexes which may be of independent interest.

1 Fréchet Distance and Persistent Homology

The Fréchet distance is a popular way to define the distance between curves. If we identify curves with continuous maps from the interval $[0, 1]$ to the plane (for now), then the Fréchet distance between $f, g : [0, 1] \rightarrow \mathbb{R}^2$ is defined as

$$\mathbf{d}_F(f, g) := \inf_{h \in H} \sup_{t \in [0, 1]} \|f(t) - g(h(t))\|,$$

where H is the set of orientation-preserving homeomorphisms from $[0, 1]$ to itself. Composition with a homeomorphism h allows one to ignore differences that only appear because of a particular parameterization of the curves.

Other variants include the homotopic [6] and nonmonotone [2] Fréchet distance. Efficient approximation algorithms are also known [15].

The generalized Fréchet distance can be defined for other maps from (compact) topological spaces into metric spaces. That is, for $f, g : X \rightarrow Y$,

$$\mathbf{d}_F(f, g) := \inf_{h \in H} \sup_{t \in X} \mathbf{d}_Y(f(t), g(h(t))),$$

where H is the set of homeomorphisms $X \rightarrow X$. This has also been studied for simple polygons [4], folded polygons [18], and other surfaces [20, 5]. It should be noted that the general results for surfaces require exponential time. Even for maps from $\mathbb{R}^2 \rightarrow \mathbb{R}$, the problem was only just recently shown to be in NP and is NP-Hard to approximate within $2 - \varepsilon$ [5].

*This work was partially supported under grants CCF-1652218 and CCF-1525978.

In this paper, we propose a new way to bound the generalized Fréchet distance using persistent homology. The great strength of this approach (or weakness depending on your perspective) is that it provides meaningful bounds without searching for the explicit homeomorphism. Instead, it computes a signature that is invariant to homeomorphisms. It simply sees past the parameterization. Our approach is very similar to both the persistent homology transform of Turner et al. [26] and the persistence distortion distance of Dey et al. [14]. However, we extend these ideas to arbitrary metrics and give the first connections between these approaches and the Fréchet distance (see Section 4). Previous work that used persistence to classify trajectories operated on the range rather than the domain of the trajectories [24, 23] which results in a (usually much) higher dimensional computation. Persistence diagrams have also been used for computing distances between road networks [1]. Another important related line involves the natural pseudodistance from size theory, which is equivalent to the generalized Fréchet distance for functions into \mathbb{R}^d [16, 3].

The primary motivation for considering lower bounds on the Fréchet distance is that they provide a fast way to filter far apart trajectories or functions. Thus, given a data set of many functions, one can avoid lengthy Fréchet distance computations if the distance between their persistence signatures is large. This would be useful, for example, when clustering functions, or computing representative samples. In such cases, it is not important to know the minimizing homeomorphism, because a certificate that the distance is larger than some constant suffices.

We consider the most common domains for this approach will be simplicial complexes reconstructed from point clouds. In this case, we show in Section 5 that the technique can still be used when the domains are not homeomorphic, but only ε -homotopy equivalent (see Section 2 for the definition). This is the type of guarantee provided by several reconstruction results. Thus, we can still compute a bound on the Fréchet distance between functions from discrete approximations, even if we don't have homeomorphic reconstructions.

In Section 6, we resolve the algorithmic challenge of computing exactly these persistence signatures for functions from piecewise linear simplicial complexes to convex metric spaces. As we illustrate there, the naive approach can be very sensitive to parameterization, exactly the problem we wish to avoid.

2 Background

Functions. For functions $f : A \rightarrow B$ and $g : B \rightarrow C$, we let $g \circ f$ denote the *composition* $(g \circ f)(a) = g(f(a))$. For a subset $A' \subseteq A$, we let $f|_{A'}$ denote the *restriction* of f to the domain A' . Real-valued functions over a common domain have a natural partial order where we write $f \leq g$ to indicate that $f(x) \leq g(x)$ for all x in the domain. If the domain A of f is a metric space, we say that f is *1-Lipschitz* if for all $a, b \in A$, we have $f(a) \leq f(b) + \mathbf{d}(a, b)$. We will abuse notation slightly and identify real numbers with constant-valued functions when the domain is clear. For example, if $\varepsilon \in \mathbb{R}$, then $f \leq \varepsilon$ means that $f(x) \leq \varepsilon$ for all x in the domain.

Topological Basics. A *homeomorphism* is a continuous map with a continuous inverse. A homeomorphism gives a very strong topological equivalence between spaces. For example, a circle and a square are *homeomorphic* as are any two simple curves. Two weaker forms of equivalence are given by homotopy and homology. Two maps $f, g : X \rightarrow Y$ are *homotopic* if there is a continuous map $a : X \times [0, 1] \rightarrow Y$ such that $a(\cdot, 0) = f$ and $a(\cdot, 1) = g$. If f and g are homotopic, we write $f \sim g$. Two spaces X and Y are *homotopy equivalent* if there are continuous maps $a : X \rightarrow Y$ and

$b : Y \rightarrow X$ such that $a \circ b \sim \text{id}_Y$ and $b \circ a \sim \text{id}_X$. The spaces are ε -homotopy equivalent if moreover for each $x \in X$, and each $y \in Y$, we have

$$\mathbf{d}_X(x, b(a(x))) \leq \varepsilon \quad \text{and} \quad \mathbf{d}_Y(y, a(b(y))) \leq \varepsilon.$$

We write $X \simeq Y$ if X and Y are homotopy equivalent and $X \simeq_\varepsilon Y$ if X and Y are ε -homotopy equivalent. As examples, a circle and an annulus are homotopy equivalent even though they are not homeomorphic. The circle and the annulus are ε -homotopy equivalent if the annulus is the set of points within distance ε of the circle.

Homology gives a yet weaker notion of equivalence, where “weaker” means more things are equivalent. Its somewhat more technical algebraic definition is offset by the usefulness afforded by being computable. In computational geometry and topological data analysis, one often considers simplicial homology over a field. In that setting, the homology of a simplicial complex is a vector space whose dimension counts interesting topological features such as connected components and holes. We refer the reader to standard references for the technical definition [19].

Persistent Homology. For a function $f : X \rightarrow \mathbb{R}$, where X is a topological space, we can define a nested sequence of spaces $(X_r)_{r \in \mathbb{R}}$, where

$$X_r := \{x \in X \mid f(x) \leq r\}.$$

This sequence is called a *filtration*. We will often refer to the pair (X, f) as a *filtered space*. The *persistent homology* of f describes the homology of the inclusion map $X_r \hookrightarrow X_s$ for all $r \leq s$. The sequence of homology groups $(\mathbf{H}(X_r))_{r \geq 0}$ and the maps $(\mathbf{H}(X_r \hookrightarrow X_s))_{s \geq r \geq 0}$ form the *persistence module* of the filtered space (X, f) . Persistent homology tracks the appearance (birth) and disappearance (death) of topological features as the space changes from $X_{-\infty}$ to X_∞ . Throughout we will assume our functions are *tame* in that the homology of X_r is always finite-dimensional.

A homomorphism between two persistence modules is a collection of maps between the homology groups at the same index that commute with the corresponding maps at those indices. That is, for persistence modules $\mathbb{X} = ((\mathbf{H}(X_r)), (i_r^s : \mathbf{H}(X_r) \rightarrow \mathbf{H}(X_s)))$ and $\mathbb{Y} = ((\mathbf{H}(Y_r)), (j_r^s : \mathbf{H}(Y_r) \rightarrow \mathbf{H}(Y_s)))$, we would have homomorphisms $\phi_r : \mathbf{H}(X_r) \rightarrow \mathbf{H}(Y_r)$ for all $r \geq 0$ such that $\phi_s \circ i_r^s = j_r^s \circ \phi_r$. The persistence modules are isomorphic if the homomorphism at each index is an isomorphism.

Every persistence module \mathbb{X} has a natural homomorphism into itself by shifting indices. For $\varepsilon \geq 0$, the ε -shift is denoted $\mathbb{1}_\mathbb{X}^\varepsilon$. Two persistence modules \mathbb{X} and \mathbb{Y} are ε -interleaved if there exist homomorphisms $\Psi : \mathbb{X} \rightarrow \mathbb{1}_\mathbb{Y}^\varepsilon$ and $\Phi : \mathbb{Y} \rightarrow \mathbb{1}_\mathbb{X}^\varepsilon$ such that $\Phi \circ \Psi = \mathbb{1}_\mathbb{X}^{2\varepsilon}$ and $\Psi \circ \Phi = \mathbb{1}_\mathbb{Y}^{2\varepsilon}$. The smallest ε for which \mathbb{X} and \mathbb{Y} are ε -interleaved is called the *interleaving distance*.

Homeomorphism Invariance The persistent homology of filtered spaces is invariant to homeomorphism in the following sense. The proof of this observation (see [11, 10]) is included for completeness.

Theorem 2.1. (HOMEOMORPHISM INVARIANCE OF PERSISTENT HOMOLOGY) *Let X be a topological space and let $h : X \rightarrow X$ be a homeomorphism. If $f : X \rightarrow \mathbb{R}_{\geq 0}$ is a tame, continuous function, then $f \circ h$ is tame and $\text{Pers}(f) = \text{Pers}(f \circ h)$.*

Proof. For any $\alpha \geq 0$, the set $F_\alpha = f^{-1}[0, \alpha]$ is homeomorphic to $G_\alpha = (f \circ h)^{-1}[0, \alpha]$. The desired homeomorphism is just the restriction h_α of h to F_α . It follows from the definitions that $h_\alpha(x) \in F_\alpha$

if and only if $x \in G_\alpha$ and similarly for the inverse. Let $\beta \geq \alpha$ be fixed and let i and j denote the inclusions $i : F_\alpha \rightarrow F_\beta$ and $j : G_\alpha \rightarrow G_\beta$. Then, $h_\beta \circ i = j \circ h_\alpha$. Thus, (h_α) induces an isomorphism of persistence modules, and so, $\text{Pers}(f) = \text{Pers}(f \circ h)$. \square

Persistence Barcodes. The isomorphism class of a persistence module has a nice representation as a set of pairs of numbers (b, d) where b is the *birth* time and d is the *death* time of a homological feature. This information is often depicted in the *persistence barcode*, where the pairs (b, d) are drawn as intervals or “bars”.

A feature that dies almost immediately after being born may often be ignored. So, for a barcode D , we write $D_\varepsilon := \{(b, d) \in D \mid d - b > \varepsilon\}$. This “filters” out the least persistent bars.

An ε -*matching* between two persistence barcodes A, B is a bijection $m : A' \rightarrow B'$, where $A_\varepsilon \subseteq A' \subseteq A$, $B_\varepsilon \subseteq B' \subseteq B$, and for all $a \in A'$, we have $\|a - m(a)\|_\infty \leq \varepsilon$. That is, we can match points in A with nearby points in B , possibly ignoring some points with low persistence.

The *bottleneck distance* between two persistence barcodes A and B is

$$\mathbf{d}_B(A, B) := \min\{\varepsilon \mid \text{an } \varepsilon\text{-matching } m : A \rightarrow B \text{ exists}\}$$

The bottleneck distance can be used to characterize the stability of persistence barcodes with respect to small changes in a function (see Chazal et al. [7] for a complete treatment of stability).

Theorem 2.2 (Stability of Persistence). *Let X be a topological space. For any tame functions $f, g : X \rightarrow \mathbb{R}$,*

$$\mathbf{d}_B(\text{Pers}(f), \text{Pers}(g)) \leq \|f - g\|_\infty$$

This is a special case of a more general theorem relating bottleneck distance and interleaving of persistence modules (again, see [7]).

Theorem 2.3. *If \mathbb{X} and \mathbb{Y} are ε -interleaved persistence modules, then $\mathbf{d}_B(\mathbb{X}, \mathbb{Y}) \leq \varepsilon$.*

This last theorem generalizes the fact that isomorphic persistence modules have identical persistence barcodes. It allows us to work primarily with interleavings of persistence module. However, the results are easiest to interpret and visualize with respect to the bottleneck distance.

3 Two Perspectives on Persistent Homology

Most popular presentations of persistent homology emphasize shape. Usually there is a distance function, maybe a point cloud, and invariably, a picture of a torus. Such a treatment is complemented by various stability theorems which imply, for example that for compact subsets A and B of a metric space, the distance functions $a(x) := \min_{y \in A} \mathbf{d}(x, y)$ and $b(x) := \min_{y \in B} \mathbf{d}(x, y)$ satisfy the following inequality.

$$\mathbf{d}_B(\text{Pers}(a), \text{Pers}(b)) \leq \mathbf{d}_H(A, B).$$

This is easy to visualize as the right hand side is in terms of Hausdorff distance, which intuitively corresponds to the amount you need to squint so that you can't see the difference between the sets. If A and B are respectively the images of continuous maps $f_A, f_B : X \rightarrow Y$, then

$$\mathbf{d}_H(A, B) \leq \mathbf{d}_F(f_A, f_B).$$

This follows rather directly from the definitions of the two metrics. Combining the preceding two inequalities, gives that $\mathbf{d}_B(\text{Pers}(a), \text{Pers}(b)) \leq \mathbf{d}_F(f_A, F_B)$, which *is* a relationship between bottleneck distance and Fréchet distance, but it is rather weak in that it passes through the Hausdorff distance. Indeed, the failure of the Hausdorff distance to discriminate between curves motivates the Fréchet distance in the first place.

For example, consider the unit circle. It is the image of both the maps $f(t) = (\cos(2\pi t), \sin(2\pi t))$ and $g(t) = (\cos(4\pi t), \sin(4\pi t))$. So, just looking at the images (i.e. with the Hausdorff distance), one cannot see a difference. However, their Fréchet distance is nonzero, because it is sensitive to the fact that one curve traverses the circle twice and the other only once. The distance we will define in the next section will be similarly sensitive to such differences.

From the shape-oriented perspective, one decides if persistent homology is the right tool based on whether the problem involves shapes. Then, one replaces shapes with distance functions and uses the persistent homology of the distance functions to *see past* geometrically small aberrations. This is often a valid approach, but as we see, it doesn't give new insights for bounding the Fréchet distance.

The alternative perspective on persistent homology espoused in this paper emphasizes the topological aspects over the geometric aspects, particularly the invariance of persistent homology under homeomorphism. Recall this means that $\text{Pers}(g) = \text{Pers}(g \circ h)$ where $g : X \rightarrow \mathbb{R}$ is tame and $h : X \rightarrow X$ is any homeomorphism. For a curve, a change in the parameterization is such a homeomorphism. The Fréchet distance has a similar invariance to homeomorphism in that

$$\mathbf{d}_F(f, g) = \mathbf{d}_F(f, g \circ h).$$

This follows from the fact that the composition of two homeomorphisms is also a homeomorphism, so

$$\begin{aligned} \mathbf{d}_F(f, g) &= \inf_{h_0 \in H} \sup_{t \in X} \mathbf{d}(f(t), g(h_0(t))) \\ &= \inf_{h_1 = h^{-1} \circ h_0 \in H} \sup_{t \in X} \mathbf{d}(f(t), (g \circ h)(h_1(t))) \\ &= \mathbf{d}_F(f, g \circ h). \end{aligned}$$

The critical difference is that we are looking at homeomorphism of the domain rather than the images and we are not worried about limiting ourselves to geometrically small transformations. We believe that this perspective will find more uses in future work.

4 Fréchet-stable Signatures

The persistence stability theorem says that the mapping from functions to persistence barcodes is 1-Lipschitz as a map between metric spaces. We will use the bottleneck distance as well as the homeomorphism invariance and stability of persistence barcodes to prove the following relationship between bottleneck distance and Fréchet distance.

Theorem 4.1. *If $f, g : [0, 1] \rightarrow \mathbb{R}^2$ are two curves in the plane, and $n(y) = \|y\|$ is the Euclidean norm, then*

$$\mathbf{d}_B(\text{Pers}(n \circ f), \text{Pers}(n \circ g)) \leq \mathbf{d}_F(f, g).$$

The preceding theorem is actually a corollary of the more general Theorem 4.2 involving the generalized Fréchet distance and a so-called basis defined as follows. Given a set of points $P \subset Y$, we define for each $p \in P$, the function

$$n_p(y) := \mathbf{d}_Y(y, p).$$

We call the points $p \in P$, *basis points*, and the functions n_p , *basis functions*. Then, for $f : X \rightarrow Y$, we define the *persistence signature of f with respect to P* as the following set of persistence barcodes.

$$\text{Sig}(f, P) := \{\text{Pers}(n_p \circ f) \mid p \in P\}$$

The bottleneck distance can be used to define a metric from these signatures for a given basis P , where

$$\mathbf{d}_{\text{Sig}}(f, g, P) := \max_{p \in P} \mathbf{d}_B(\text{Pers}(n_p \circ f), \text{Pers}(n_p \circ g)).$$

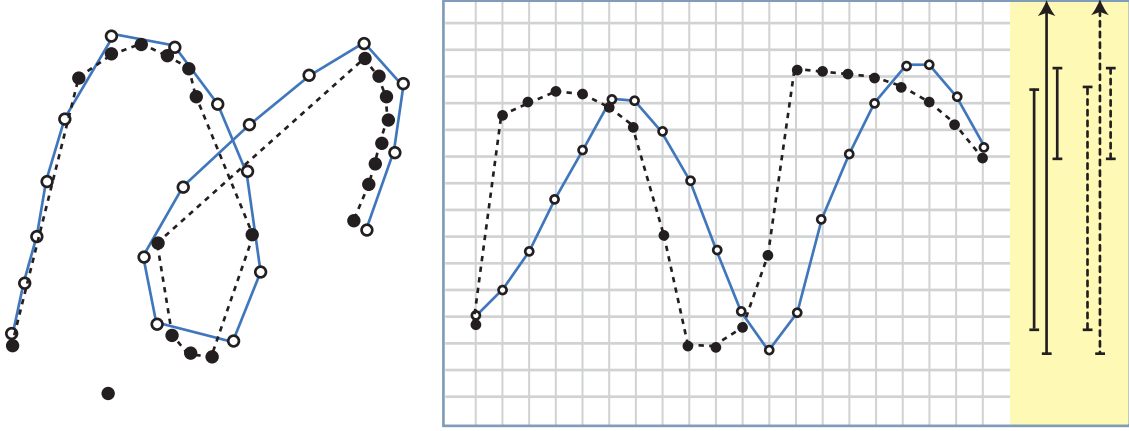


Figure 1: Two curves are shown on the left. On the right are their distance functions to a point. Different parameterizations can lead to quite different distance functions. However, the barcodes (far right) remain similar.

We are now ready to state and prove the main theorem.

Theorem 4.2. *Let X be a topological space and let (Y, \mathbf{d}_Y) be a metric space. For any two continuous functions $f, g : X \rightarrow Y$ and any finite $P \subset Y$, we have that*

$$\mathbf{d}_{\text{Sig}}(f, g, P) \leq \mathbf{d}_F(f, g).$$

Before proceeding to the proof, we observe that Theorem 4.1 follows from Theorem 4.2 by letting $X = [0, 1]$, $Y = \mathbb{R}^2$, and $P = \{0\}$.

Proof. Fix any $f, g : X \rightarrow Y$ and $P \subset Y$. Let $h : X \rightarrow X$ be any homeomorphism and let $p \in P$

be any point.

$$\begin{aligned}
& \mathbf{d}_B(\text{Pers}(n_p \circ f), \text{Pers}(n_p \circ g)) \\
&= \mathbf{d}_B(\text{Pers}(n_p \circ f), \text{Pers}(n_p \circ g \circ h)) && \text{[by Thm. 2.1]} \\
&\leq \| (n_p \circ f) - (n_p \circ g \circ h) \|_\infty && \text{[by Thm. 2.2]} \\
&= \sup_{x \in X} |\mathbf{d}_Y(f(x), p) - \mathbf{d}_Y(g(h(x)), p)| && \text{[by definition]} \\
&\leq \sup_{x \in X} \mathbf{d}_Y(f(x), g(h(x))) && \text{[triangle ineq.]}
\end{aligned}$$

Because the above inequality holds for all homeomorphisms h , it holds for the infimum and so

$$\mathbf{d}_B(\text{Pers}(n_p \circ f), \text{Pers}(n_p \circ g)) \leq \mathbf{d}_F(f, g).$$

Moreover, because this holds for all $p \in P$, it follows that $\mathbf{d}_{\text{Sig}}(f, g, P) \leq \mathbf{d}_F(f, g)$. \square

Note that the proof implies that one can replace the specific choice of the basis functions n_p with any other Lipschitz function on Y . For example, one could compute the minimum distance to the entire set P rather than individual points. This would introduce unnecessary and counterproductive symmetries into the signatures as the function would obscure which was the nearest point at any given time. The specific choice of basis functions was intended to be an example set that exists in any metric space.

5 Persistent Homology as a Topological Invariant of Functions

The definition of the Fréchet distance doesn't allow for comparison between functions defined on different (i.e. non-homeomorphic) spaces. However, there are natural cases in which one would like to do such a comparison, especially when the images of the functions are reconstructions from a point cloud. Indeed, this is a natural case to consider as there are several results that allow one to reconstruct a simplicial complex that is homotopy equivalent, though possibly not homeomorphic to an unknown underlying space. Some notable examples include the work of Niyogi, Smale, and Weinberger [22, 21] and the extension by Chazal and Lieutier [8]. Also, techniques related to the Flow Complex of Giesen and John [17] such as those developed by Dey et al. [13] and Sadri [25] also give such topological guarantees.

It is currently unknown how to even define the Fréchet distance for functions defined on non-homeomorphic domains, however, the persistence signature distance can be computed. In this section, we show that the types of guarantees produced by many of the reconstruction results listed above can also guarantee a bound on the signature distance. This means that:

If one wants to bound the Fréchet distance between two unknown functions $f, g : X \rightarrow Y$, it suffices to bound $\mathbf{d}_{\text{Sig}}(\hat{f}, \hat{g}, P)$ for “sufficiently good” reconstructions \hat{f} and \hat{g} and any set $P \subset Y$.

What exactly counts as “sufficiently good” will be defined below, but we start with a minimum condition. If $f : X \rightarrow Y$, $\hat{f} : \hat{X} \rightarrow Y$ and $H_*(X) \not\cong H_*(\hat{X})$, then the $\mathbf{d}_{\text{Sig}}(f, \hat{f}, P)$ will be unbounded. The reason is simply that nontrivial cycles in the homology produce infinite bars in the barcode and if the homology differs, then at least one will be left unmatched by any ε -matching. Even if the spaces are homotopy equivalent, there is no guarantee that the homotopy equivalence is nice with respect to the metric.

Homotopy equivalence is generally a more restrictive equivalence than homology and one might expect that the homeomorphism of the previous section might be replaced by a homotopy equivalence. However, this is false. Too much information is lost in a homotopy equivalence between filtered spaces.

Let (X, f) and (Y, g) be filtered spaces. Let \mathbb{X} and \mathbb{Y} be the corresponding persistence modules. We are first interested in giving conditions for which a map $h : Y \rightarrow X$ induces a homomorphism of the persistence modules $\mathbb{Y} \rightarrow \mathbb{X}$ or a shifting homomorphism $\mathbb{Y} \rightarrow \mathbb{1}^\varepsilon(\mathbb{X})$ for some small $\varepsilon \in \mathbb{R}$. For $\alpha \in \mathbb{R}$, let $h_\alpha := h|_{X_\alpha}$.

Lemma 5.1. *If $f \circ h \leq g + \varepsilon$, then h induces a homomorphism $\mathbb{Y} \rightarrow \mathbb{1}^\varepsilon(\mathbb{X})$.*

Proof. The homomorphism will be composed of the maps induced on homology by the maps h_α . It will suffice to show that the maps exist and commute appropriately on the filtered spaces. We first need to show that $h_\alpha(Y_\alpha) \subseteq X_{\alpha+\varepsilon}$. Fix any $y \in Y_\alpha$. So, $\alpha \geq g(y) \geq f(h(y)) + \varepsilon = f(h_\alpha(y)) + \varepsilon$. So, it follows that $h_\alpha(y) \in X_{\alpha+\varepsilon}$. To complete the proof, it suffices to observe that restrictions h_α and h_β commute with the inclusions $Y_\alpha \hookrightarrow Y_\beta$ and $X_{\alpha+\varepsilon} \hookrightarrow X_{\beta+\varepsilon}$. Thus, the induced maps also commute in homology. \square

Now, suppose $h : Y \rightarrow X$ is a homotopy equivalence. Let \bar{h} denote its homotopy inverse. Let $a_X : X \times [0, 1] \rightarrow X$ and $a_Y : Y \times [0, 1] \rightarrow Y$ be homotopies showing that $h \circ \bar{h} \sim \text{id}_X$ and $\bar{h} \circ h \sim \text{id}_Y$ respectively.

Definition 5.1. *The homotopy equivalence $h : Y \rightarrow X$ is ε -monotone with respect to $f : X \rightarrow \mathbb{R}$ and $g : Y \rightarrow \mathbb{R}$ if*

- (i) $f \circ h \leq g + \varepsilon$ and $g \circ \bar{h} \leq f + \varepsilon$.
- (ii) For all $t \in [0, 1]$: $f \circ a_X(\cdot, t) \leq f + 2\varepsilon$ and $g \circ a_Y(\cdot, t) \leq g + 2\varepsilon$.

The preceding definition gives a sufficient condition for the persistent homology to be approximately preserved by the homotopy equivalence.

Lemma 5.2. *If $h : Y \rightarrow X$ is an ε -monotone homotopy equivalence with respect to $f : X \rightarrow \mathbb{R}$ and $g : Y \rightarrow \mathbb{R}$, then $\text{Pers}(f)$ and $\text{Pers}(g)$ are ε -interleaved.*

Proof. Condition (i) of the definition of ε -monotone and Lemma 5.1 imply that h and \bar{h} induce homomorphisms $\Phi : \mathbb{Y} \rightarrow \mathbb{1}_X^\varepsilon(\mathbb{X})$ and $\Psi : \mathbb{X} \rightarrow \mathbb{1}_Y^\varepsilon(\mathbb{Y})$ respectively. Condition (ii) in the definition of ε -monotone implies the interleaving condition $\Phi \circ \Psi = \mathbb{1}_X^{2\varepsilon}$ and $\Psi \circ \Phi = \mathbb{1}_Y^{2\varepsilon}$ because the homomorphisms are all induced by homotopic maps and are therefore equal in homology. \square

Definition 5.2. *Let X and Y be subsets of a metric space M . A homotopy equivalence $h : Y \rightarrow X$ is ε -metric preserving if for all $x \in X$ and all $y \in Y$,*

- (i) $\mathbf{d}(y, h(y)) \leq \varepsilon$ and $\mathbf{d}(x, \bar{h}(x)) \leq \varepsilon$.
- (ii) For all $t \in [0, 1]$, $\mathbf{d}(x, a_X(x, t)) \leq 2\varepsilon$ and $\mathbf{d}(y, a_Y(y, t)) \leq 2\varepsilon$.

Lemma 5.3. *If $f : M \rightarrow \mathbb{R}$ is 1-Lipschitz and $h : Y \rightarrow X$ is an ε -metric-preserving homotopy equivalence, then $\text{Pers}(f|_X)$ and $\text{Pers}(f|_Y)$ are ε -interleaved.*

Proof. The conditions in the definition of ε -metric preserving translate exactly into the conditions in the definition of ε -monotone when you apply the Lipschitz condition. The result then follows from Lemma 5.2. \square

We can now prove a general theorem that gives a bound on the signature distance between two functions defined on homotopy equivalent domains.

Theorem 5.4. *Let X and Y be subsets of a metric space M . Let $h : Y \rightarrow X$ be an ε -metric preserving homotopy equivalence. For the inclusion maps $f : X \hookrightarrow M$ and $g : Y \hookrightarrow M$, we have $\mathbf{d}_{\text{Sig}}(f, g, P) \leq \varepsilon$ for all finite $P \subset M$.*

Proof. For each basis point $p \in P$, the basis function n_p is 1-Lipschitz. Observe that for inclusion maps, $n_p|_{\text{im}f} = n_p \circ f$. The result now follows from the definition of the signature distance and Lemma 5.3. \square

It may seem like an excessive restriction to only consider inclusion maps, but simple examples show that some similarly strong constraint is necessary to guarantee that function values of close points stay close. It is possible to combine Theorem 4.2 and Theorem 5.4 by considering reparameterizations of X and Y that are homeomorphisms.

6 Discretizing Convex Functions on Piecewise Linear Simplicial Complexes

The simplicity of the proof of Theorem 4.2 belies the fact that it is not clear how to compute these signatures in general. Usually, persistent homology is computed using some piecewise linear function over some discretization of the domain such as a simplicial complex. However, the functions defined in our signatures do not necessarily come with such a structure, *even if the domain is already a simplicial complex*. This is not surprising given the generality of the result.

In this section, we will show how this computation can be performed in the most common setting, piecewise-linear simplicial complexes in \mathbb{R}^d . Our proof will consider the slightly more general case of arbitrary convex functions over piecewise-linear simplicial complexes.

6.1 What could go wrong?

Let us start with a very simple example to illustrate the difficulty. Suppose we want to bound the Fréchet distance between two curves, each with just two straight segments. Given, the preceding sections, it might seem reasonable to just compare the persistence barcodes of the simplicial filtrations induced by the distance function to the origin. Recall that an *induced filtration*, also known as a *lower-star filtration*, is equivalent to the piecewise linear (PL) extension of a function on the vertices to the rest of the simplicial complex. It is a common practice in TDA to rely on such PL-approximations to give a simplicial filtration in a format expected by most persistent homology codes. Also, it is easy to compute as it only requires computing function values at the vertices. However, in this case, the naive use of a PL-approximation can lead to a very bad answer. In this section, we will show how the naive approach can fail and how to fix it.

In the figure below, we have two curves that are close in Fréchet distance, but they are split into two pieces in different ways. The distance functions (as shown on the left of the figure) of each of these curves to the origin are correspondingly close and as are their persistence barcodes. However,

the induced distance functions (shown in the middle) and the resulting difference in persistence are significant. By splitting any segment that achieves a local minimum in its relative interior, we can produce an induced filtration (shown on the right) with exactly the same persistence barcode as the true distance function (as shown on the left). In this section, we show how to perform an equivalent subdivision for an arbitrary PL simplicial complex and piecewise convex (distance) function. The main result is an effective means of producing PL filtrations for the kind of distance functions used for the signatures in the preceding sections.

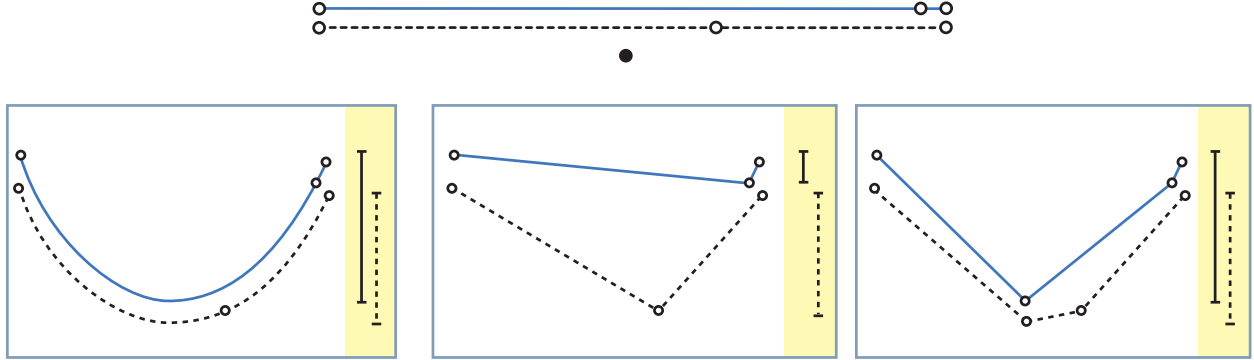


Figure 2: (Left) The distance functions to a point are shown for each of the curves on top. (Center) The PL distance function induced on the given segments. (Right) The PL distance function induced on a subdivision of the given segments. The barcodes on the left and right are identical, whereas the barcode in the middle can be quite different.

6.2 An Equivalent PL Filtration

We first show that any convex function on a simplicial complex for which the minimum value on each simplex is achieved at a vertex has the same persistent homology as the induced filtration on the same complex. The proof of this fact is a straightforward application of the Persistent Nerve Lemma (PNL) [9]. Let $U = \{\{U_1^\alpha\}_{\alpha \geq 0}, \dots, \{U_n^\alpha\}_{\alpha \geq 0}\}$ be a collection of filtrations. The nerve of U is the filtered simplicial complex $\text{Nrv } U$ in which the complex at scale α is $\{\sigma \subset [n] \mid \bigcap_{i \in \sigma} U_i^\alpha \neq \emptyset\}$. The PNL implies that if U_i^α is convex¹ for all i , then the nerve of U has the same persistent homology as the union filtration $\{\bigcup_{i \in [n]} U_i^\alpha\}_{\alpha \geq 0}$.

Lemma 6.1. *Let K be a finite simplicial complex. Let $f : \underline{K} \rightarrow \mathbb{R}$ be a function such that for all $\sigma \in K$, we have $f|_{\underline{\sigma}}$ is convex and achieves its minimum value at a vertex. Let $\hat{f} : \underline{K} \rightarrow \mathbb{R}$ be the PL filtration on \underline{K} induced by f . Then, $\text{Pers}(f) = \text{Pers}(\hat{f})$.*

Proof. For each simplex $\sigma \in K$ and each $\alpha \in \mathbb{R}$, let $U_\sigma^\alpha := \{x \in \underline{\sigma} \mid f(x) \leq \alpha\}$. Similarly, let $\hat{U}_\sigma^\alpha := \{x \in \underline{\sigma} \mid \hat{f}(x) \leq \alpha\}$. Note that the sets U_σ^α and \hat{U}_σ^α are convex. Let $U^\alpha = \{U_\sigma^\alpha \mid \sigma \in K\}$ and let $\hat{U}^\alpha = \{\hat{U}_\sigma^\alpha \mid \sigma \in K\}$. For all $\alpha \in \mathbb{R}$, the collection U^α and \hat{U}^α form good covers of the α -sublevel sets of f and \hat{f} respectively, so that by the Persistent Nerve Lemma, $\text{Pers}(f) = \text{Pers}(\{\text{Nrv } U^\alpha\}_{\alpha \geq 0})$ and $\text{Pers}(\hat{f}) = \text{Pers}(\{\text{Nrv } \hat{U}^\alpha\}_{\alpha \geq 0})$. It will suffice to show that for any α , we have $\text{Nrv } U^\alpha =$

¹The exact condition is that the sets should be open and all intersections should be empty or contractible. It is easy to check that for Hausdorff metric topologies, being convex suffices as the open metric neighborhoods of the sets produce the same nerve.

Nrv \hat{U}^α , i.e., they are the same filtered simplicial complex. This follows from the assumption that the function f is minimized at a vertex for each simplex and, by definition, f and \hat{f} agree at the vertices. \square

6.3 Subdividing a Simplicial Complex

We want to subdivide a simplicial complex into one that satisfies the hypothesis of Lemma 6.1. We will use the fact that convex functions take a unique minimum value on closed, convex sets and that local minima are global minima. We will need to add a single vertex for each simplex that attains its minimum value in its relative interior. For simplicity, we describe the filtration on the full barycentric subdivision. This approach could produce extra vertices.

Let K be an abstract simplicial complex, i.e. a set of vertices and a family of subsets of vertices that is closed under taking subsets. Let \underline{K} denote the geometric realization of K , for example as the image of some function $f : K \rightarrow Y$. A *flag* in K is a subset of simplices that are totally ordered by inclusion. The *abstract barycentric subdivision* of K is the simplicial complex

$$\text{bary}(K) := \{S \subset K \mid S \text{ is a flag of } K\}$$

Let p be a basis point. Let $b : \text{bary}(K) \rightarrow \underline{K}$ be defined so that $b(v) = \underline{v}$ and $b(\sigma) = \text{argmin}_{x \in \sigma} n_p(f(x))$. That is b gives a point of minimum function value on a simplex. One gets the desired subdivision of K by considering all simplices $\{b(\sigma_0), \dots, b(\sigma_k)\}$ for all flags $\sigma_0 \subseteq \sigma_k$ in $\text{bary}(K)$.

7 Future work and open questions

The bounds we give are all one sided. It is not known whether there is a good basis set for every curve or pair of curves, where “good” here means that the signature distance is a constant factor approximation to the true Fréchet distance.

It may be that a constant sized set of basis points suffice. However, it is immediate from the definitions that one ought to be able to replace the discrete set of basis points with a continuous path of basis points. In that case, the signature could be computed using the Vineyard algorithm of Cohen-Steiner et al. [12]. This would be the first step towards a coordinate-free version of the persistent homology transform of Turner et al. [26].

One could also try to use the matching of the persistence barcode to try to find a good matching between, say curves. This also doesn’t work directly as the following figure illustrates.

In a sense, the difficulty of extracting the matching from the diagram is likely the reason that the approach can avoid existing NP-hardness results for computing the Fréchet distance exactly. Still, a more nuanced approach might give a better matching.

This paper established the theoretical basis for this approach of using persistence-based signatures as a substitute for Fréchet distance. In future work, we will explore its usefulness in practice.

8 Acknowledgments

I would like to thank the anonymous reviewers for many insightful and helpful comments.

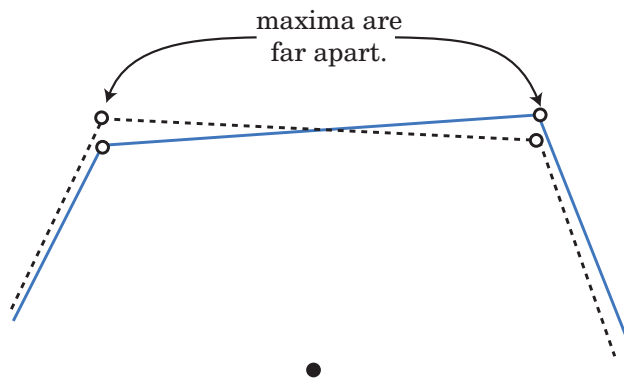


Figure 3: Two very similar curves with small signature distance and small Fréchet distance. The corresponding local maxima of the distance to the basis point can be quite far.

References

- [1] M. Ahmed, B. T. Fasy, and C. Wenk. Local persistent homology based distance between maps. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 43–52, 2014.
- [2] H. Alt and M. Godau. Computing the fréchet distance between two polygonal curves. *Int. J. Comput. Geometry Appl*, 5(1 & 2):75–91, 1995.
- [3] S. Biasotti, L. D. Floriani, B. Falcidieno, P. Frosini, D. Giorgi, C. Landi, L. Papaleo, and M. Spagnuolo. Describing shapes by geometrical-topological properties of real functions. *ACM Comput. Surv.*, 40(4):12:1–12:87, Oct. 2008.
- [4] K. Buchin, M. Buchin, and C. Wenk. Computing the fréchet distance between simple polygons. *Computational Geometry: Theory and Applications*, 41:2–20, 2008.
- [5] K. Buchin, T. Ophelders, and B. Speckmann. Computing the Fréchet distance between real-valued surfaces. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2443–2455, 2017.
- [6] E. W. Chambers, É. C. de Verdière, J. Erickson, S. Lazard, F. Lazarus, and S. Thite. Homotopic fréchet distance between curves or, walking your dog in the woods in polynomial time. *Computational Geometry*, 43(3):295 – 311, 2010.
- [7] F. Chazal, V. de Silva, M. Glisse, and S. Oudot. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics. Springer International Publishing, 2016.
- [8] F. Chazal and A. Lieutier. Smooth manifold reconstruction from noisy and non-uniform approximation with guarantees. *Computational Geometry: Theory and Applications*, 40:156–170, 2008.
- [9] 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.
- [10] D. Cohen-Steiner and H. Edelsbrunner. Inequalities for the curvature of curves and surfaces. *Foundations of Computational Mathematics*, pages 391–404, 2007.
- [11] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Comput. Geom*, 37:103–120, 2007.
- [12] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the 22nd ACM Symposium on Computational Geometry*, pages 119–126, 2006.
- [13] T. K. Dey, J. Giesen, E. A. Ramos, and B. Sadri. Critical points of distance to an ε -sampling of a surface and flow-complex-based surface reconstruction. *International Journal of Computational Geometry and Applications (IJCGA)*, 18(01n02):29–61, 2008.
- [14] T. K. Dey, D. Shi, and Y. Wang. Comparing graphs via persistence distortion. In *Proceedings of the 31st International Symposium on Computational Geometry*, pages 491–506, 2015.

- [15] A. Driemel, S. Har-Peled, and C. Wenk. Approximating the fréchet distance for realistic curves in near linear time. *Discrete & Computational Geometry*, 48(1):94–127, 2012.
- [16] P. Frosini. A distance for similarity classes of submanifolds of a euclidean space. *Bull. Austral. Math. Soc.*, 42:407–416, 1990.
- [17] J. Giesen and M. John. The flow complex: A data structure for geometric modeling. *Computational Geometry*, 39(3):178–190, 2008.
- [18] A. F. C. IV, A. Driemel, J. Sherette, and C. Wenk. Computing the fréchet distance between folded polygons. *Computational Geometry*, 50:1–16, 2015.
- [19] J. R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, 1984.
- [20] A. Nayyeri and H. Xu. On computing the fréchet distance between surfaces. In *Proceedings of the 32nd International Symposium on Computational Geometry*, pages 55:1–55:15, 2016.
- [21] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.
- [22] P. Niyogi, S. Smale, and S. Weinberger. A topological view of unsupervised learning from noisy data. *SIAM J. Comput.*, 40(3):646–663, 2011.
- [23] F. T. Pokorny, K. Goldberg, and D. Kragic. Topological trajectory clustering with relative persistent homology. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 16–23, 2016.
- [24] F. T. Pokorny, M. Hawasly, and S. Ramamoorthy. Topological trajectory classification with filtrations of simplicial complexes and persistent homology. *The International Journal of Robotics Research*, 35(1–3):204–223, 2016.
- [25] B. Sadri. Manifold homotopy via the flow complex. *Computer Graphics Forum*, 28(5):1361–1370, 2009.
- [26] K. Turner, S. Mukherjee, and D. M. Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference*, 2014.