Local Signatures using Persistence Diagrams

Mathieu Carriere, Steve Oudot, Maks Ovsjanikov

Local Signatures using Persistence Diagrams

Mathieu Carrière, Steve Y. Oudot, Maks Ovsjanikov

June 4, 2015

1 Introduction

In this article, we adress the problem of devising signatures using the framework of persistent homology. Considering a compact length space $\mathbb{X}$ with curvature bounded above, we build, either for every point $x \in \mathbb{X}$ or for $\mathbb{X}$ itself, a topological signature $V \in \mathbb{R}^d$ that is provably stable to perturbations of $\mathbb{X}$ in the Gromov-Hausdorff distance. This signature has been used in 3D shape analysis tasks, such as shape segmentation and matching [2]. Here, we provide general statements and formal proofs of stability for this signature.

2 Preliminaries

The following preliminary definitions require some background in metric geometry and persistence theory. Good introductions to these subjects can be found in [1] and [3] respectively.

2.1 Compact metric spaces

Let $\mathcal{X}$ denote the collection of all compact metric spaces. Let $(\mathbb{X}, d_{\mathbb{X}})$ and $(\mathbb{Y}, d_{\mathbb{Y}})$ be two such spaces. They are said to be *isometric* if there exists a surjective map $\phi : \mathbb{X} \to \mathbb{Y}$ that preserves distances, namely: $\forall x, x' \in \mathbb{X}$, $d_{\mathbb{Y}}(\phi(x), \phi(x')) = d_{\mathbb{X}}(x, x')$. Such a map is called an *isometry*.

Definition 2.1 *A* correspondence *between* $\mathbb{X}$ *and* $\mathbb{Y}$ *is a subset* $C \subseteq \mathbb{X} \times \mathbb{Y}$ *such that:*
– $\forall x \in \mathbb{X}$, $\exists y \in \mathbb{Y}$ *s.t.* $(x, y) \in C$,
– $\forall y \in \mathbb{Y}$, $\exists x \in \mathbb{X}$ *s.t.* $(x, y) \in C$.

Let $\mathcal{C}(\mathbb{X}, \mathbb{Y})$ denote the set of all correspondences between $\mathbb{X}$ and $\mathbb{Y}$.

Definition 2.2 *The* metric distortion *of a correspondence* $C \in \mathcal{C}(\mathbb{X}, \mathbb{Y})$ *is:*

$$\epsilon_m(C) = \sup_{(x,y),(x',y') \in C} |d_{\mathbb{X}}(x, x') - d_{\mathbb{Y}}(y, y')|$$

Definition 2.3 *The* Gromov-Hausdorff distance *between compact metric spaces* $(\mathbb{X}, d_{\mathbb{X}})$, $(\mathbb{Y}, d_{\mathbb{Y}})$ *is:*

$$\mathrm{d_{GH}}((\mathbb{X}, d_{\mathbb{X}}), (\mathbb{Y}, d_{\mathbb{Y}})) = \frac{1}{2} \inf_{C \in \mathcal{C}(\mathbb{X}, \mathbb{Y})} \epsilon_m(C)$$

The map $\mathrm{d_{GH}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ defines a metric on the set of isometry classes of compact metric spaces [1, Thm. 7.3.30].

Assume now that $\mathbb{X}, \mathbb{Y}$ are equipped with continuous real-valued functions $f : \mathbb{X} \to \mathbb{R}$ and $g : \mathbb{Y} \to \mathbb{R}$.

Definition 2.4 *The* functional distortion *of a correspondence $C \in \mathcal{C}(\mathbb{X}, \mathbb{Y})$ is defined by:*

$$\epsilon_f(C) = \sup_{(x,y)\in C} |f(x) - g(y)|$$

2.2 Length Spaces

In this article, we will make heavy use of *compact length spaces with curvature bounded above.*

Definition 2.5 *Let $\mathbb{X}$ be a topological space and I be an interval in $\mathbb{R}$. A function $c : I \to \mathbb{X}$ is called a* path *of $\mathbb{X}$.*

Definition 2.6 *Let $\mathbb{X}$ be a topological space. A* length structure *(A, L) on $\mathbb{X}$ is a set of paths A together with a map $L : A \to \mathbb{R}_+ \cup \{+\infty\}$ s.t.*

- *$\forall a \le c \le d \le b$, if $\gamma : [a, b] \to \mathbb{X} \in A$ then $\gamma|_{[c,d]} \in A$*
- *$\forall a \le c \le b$ and $\gamma : [a, b] \to \mathbb{X}$, if $\gamma|_{[a,c]} \in A$ and $\gamma|_{[c,b]} \in A$ then $\gamma \in A$ and $L(\gamma) = L(\gamma|_{[a,c]}) + L(\gamma|_{[c,b]})$*
- *Let $\gamma : [a, b] \to \mathbb{X}$ and $\phi : [c, d] \to [a, b]$ s.t. $\phi(t) = \alpha t + \beta$. Then $\gamma \circ \phi \in A$ and $L(\gamma \circ \phi) = L(\gamma)$*
- *Let $f(\cdot) = L(\gamma|_{[a,\cdot]})$. Then f is continuous.*
- *Let $x \in \mathbb{X}$ and U_x be a neighborhood of x. Then $\inf\{L(\gamma) \mid \gamma(a) = x, \gamma(b) \in \mathbb{X} - U_x\} > 0$*

Definition 2.7 *A* length space *$(\mathbb{X}, d_{\mathbb{X}})$ is a metric space endowed with a length structure (A, L) s.t. A is the set of continuous paths and*

$$\forall x, x' \in \mathbb{X}, d_{\mathbb{X}}(x, x') = \inf\ \{L(\gamma) \mid \gamma : [a, b] \to \mathbb{X} \in A \text{ and } \gamma(a) = x \quad \text{and } \gamma(b) = x'\}$$

$d_{\mathbb{X}}$ is then called a length metric.

In a length space, a path whose length is equal to the distance between its endpoints is called a *shortest path.* In a compact length space, there always exists a shortest path between any pair of points. However, it may not be unique. We also assume that all compact length spaces that we consider in this article have *finite* length metrics.

Let the k-plane be the two-dimensional model space of constant curvature k and R_k be its diameter. Let also denote the distance between two points x, x' in the k-plane by $|xx'|$.

- The k-plane is a sphere of radius $\sqrt{k}$ with its length metric if $k > 0$. One has $R_k = \frac{1}{\sqrt{k}}$.
- The k-plane is the Euclidean plane if $k = 0$. One has $R_k = +\infty$.
- The k-plane is a hyperbolic plane of curvature k if $k < 0$. One has $R_k = +\infty$.

A *k-comparison triangle* for three points a, b, c in a length space $(\mathbb{X}, d_{\mathbb{X}})$ is a triangle $(a'b'c')$ in the k-plane s.t. $|a'b'| = d_{\mathbb{X}}(a, b)$, $|b'c'| = d_{\mathbb{X}}(b, c)$, $|a'c'| = d_{\mathbb{X}}(a, c)$ and $|a'b'| + |b'c'| + |a'c'| < 2R_k$.

Definition 2.8 *A compact length space* $(\mathbb{X}, d_{\mathbb{X}})$ *has* curvature bounded above *if every point* $x \in \mathbb{X}$ *has a neighborhood* U_x *that satisfies the following:*
$\exists k \in \mathbb{R}$ *s.t. for every* $a, b, c \in U_x$ *and their* k*-comparison triangle* $(a'b'c')$*, and for every* d *in any shortest path between* a *and* c*,*

$$d_{\mathbb{X}}(b, d) \leq |b'd'|$$

where d' *is the point in* $[a'c']$ *s.t.* $|a'd'| = d_{\mathbb{X}}(a, d)$

2.3 Čech and Vietoris-Rips complexes

Basic definitions for simplicial complexes and homology can be found in Chapter I of [7]. We also recall the definition of the *Čech* and the *Vietoris-Rips* simplicial complexes. Let $P = \{p_1 \ldots p_n\}$ be a point cloud in a metric space $(\mathbb{X}, d_{\mathbb{X}})$.

Definition 2.9 $\forall \delta > 0$*, the* Čech complex $C_\delta(P, d_{\mathbb{X}})$ *is the nerve of the unions of the balls of radius* δ *that are centered on elements of* P*. More formally,* $\forall k \in \{1 \ldots n\}$*:*

$$\{p_{i_1} \ldots p_{i_k}\} \in C_\delta(P, d_{\mathbb{X}}) \Leftrightarrow \bigcap_{j=1}^{k} B_\delta(p_{i_j}, d_{\mathbb{X}}) \neq \emptyset$$

Definition 2.10 $\forall \delta > 0$*, the* Vietoris-Rips complex $R_\delta(P, d_{\mathbb{X}})$ *is defined in the following way.* $\forall k \in \{1 \ldots n\}$*:*

$$\{p_{i_1} \ldots p_{i_k}\} \in R_\delta(P, d_{\mathbb{X}}) \Leftrightarrow \max_{u,v \in \{1 \ldots k\}} d_{\mathbb{X}}(p_{i_u}, p_{i_v}) \leq \delta$$

We have the following useful relation between Čech and Vietoris-Rips complexes [5]:

$$\forall \delta \geq 0,\ C_\delta(P, d_{\mathbb{X}}) \subseteq R_{2\delta}(P, d_{\mathbb{X}}) \subseteq C_{2\delta}(P, d_{\mathbb{X}}). \tag{1}$$

2.4 Nested Persistence Modules

We use the notion of *nested persistence modules.*

Definition 2.11 *Let* $G = \{G_\alpha\}_{\alpha \in \mathbb{R}}$ *and* $G' = \{G'_\alpha\}_{\alpha \in \mathbb{R}}$ *be two filtrations such that* $\forall \alpha \in \mathbb{R}, G_\alpha \subseteq G'_\alpha$*. The nested persistence module* $\mathbb{U}_{G \hookrightarrow G'}$ *is the image of the morphism between persistence modules* $\{H_p(G_\alpha)\}_{\alpha \in \mathbb{R}}$ *and* $\{H_p(G'_\alpha)_{\alpha \in \mathbb{R}}\}$ *induced at homology level by the canonical inclusion* $G_\alpha \hookrightarrow G'_\alpha$*. More precisely,*

$$(\mathbb{U}_{G \hookrightarrow G'})_\alpha = Im(g_\alpha)$$

where g_α *is the morphism induced by the inclusion map* $H_p(G_\alpha) \to H_p(G'_\alpha)$*.*

Given an increasing (w.r.t. inclusion) sequence $\{P_\alpha\}_{\alpha \in \mathbb{R}}$ of subsets of a finite metric space (P, d_P) and Rips parameters $0 \leq \delta \leq \delta'$, we write $\{R_{\delta \to \delta'}(P_\alpha, d_P)\}_{\alpha \in \mathbb{R}}$ to denote the persistence module of the nested pair of Vietoris-Rips filtrations $\{R_\delta(P_\alpha, d_P) \hookrightarrow R_{\delta'}(P_\alpha, d_P)\}_{\alpha \in \mathbb{R}}$.

3 Stability Theorems

3.1 Persistence diagrams

We start with the following lemma from [4]:

Lemma 3.1 (from [4]) *Let $(\mathbb{X}, d_{\mathbb{X}})$ be a compact length space of curvature bounded above, and let $Q \subseteq \mathbb{X}$ be a finite ε-sample of $\mathbb{X}$. Let also $f : \mathbb{X} \to \mathbb{R}$ and $g : Q \to \mathbb{R}$ be two maps such that f is c-Lipschitz. Suppose there exist $\varepsilon', \varepsilon'' \in [\varepsilon, \varrho(\mathbb{X}))$ and two filtrations $\{G_\alpha\}_{\alpha\in\mathbb{R}}$ and $\{G'_\alpha\}_{\alpha\in\mathbb{R}}$ such that for all $\alpha \in \mathbb{R}$ we have:*

$$C_\varepsilon(g^{-1}((-\infty, \alpha]), d_{\mathbb{X}}) \subseteq G_\alpha \subseteq C_{\varepsilon'}(g^{-1}((-\infty, \alpha]), d_{\mathbb{X}}) \subseteq G'_\alpha \subseteq C_{\varepsilon''}(g^{-1}((-\infty, \alpha]), d_{\mathbb{X}}). \tag{2}$$

Then,

$$\mathrm{d}_{\mathrm{b}}^{\infty}(\mathrm{PD}(f), \mathrm{PD}(\mathbb{U}_{G\hookrightarrow G'})) \leq c\varepsilon'' + \max_{q\in Q} |f(q) - g(q)|$$

Let us point out that the above lemma is in fact a slight variant of Lemma 5 of [4], whose proof is the same except for the evocation of Lemma 6 instead of Lemma 1 in Eq. (6) of that paper. Note also that the result is stated for compact Riemannian manifolds in [4], however, as mentioned in Section 2.1 of that paper, a close look at the proof reveals that the Riemannian structure itself is not exploited, only the fact that small enough open metric balls and their intersections are contractible, so the Nerve Lemma [6, Corollary 4G.3] and its persistent version [5, Lemma 3.4] can be applied to unions of (small enough) balls. Now, compact length spaces $(\mathbb{X}, d_{\mathbb{X}})$ of curvature bounded above have a positive *convexity radius* $\varrho(\mathbb{X})$, such that any open metric ball of radius less than $\varrho(\mathbb{X})$ is convex in the sense that any two points x, y in such a ball are connected by a unique shortest path, and that this path is contained within the ball and depends continuously on the positions of x and y — see e.g. Propositions 9.1.16 and 9.1.17 in [1]. As a result, open metric balls of radius less than $\varrho(\mathbb{X})$, as well as their intersections, are contractible. This makes it possible for us to rephrase the result from [4] in the more general context of compact length spaces of curvature bounded above.

Lemma 3.2 *Let $(\mathbb{X}, d_{\mathbb{X}})$ be a compact length space of curvature bounded above. Let (P, d_P) be a finite metric space, and let $f : \mathbb{X} \to \mathbb{R}$ and $g : P \to \mathbb{R}$ such that f is c-Lipschitz. Let $P_\alpha = g^{-1}((-\infty, \alpha])$. Assume that $\mathrm{d}_{\mathrm{GH}}(P, \mathbb{X}) < \frac{\varrho(\mathbb{X})}{20}$. Then, for any correspondence $C \in \mathcal{C}(P, \mathbb{X})$ such that $\varepsilon_m(C) < \frac{\varrho(\mathbb{X})}{10}$, and for any parameters $\delta \in (3\varepsilon_m(C), \frac{\varrho(\mathbb{X})}{2} - 2\varepsilon_m(C))$ and $\delta' \in (2\delta + 3\varepsilon_m(C), \varrho(\mathbb{X}) - \varepsilon_m(C))$,*

$$\mathrm{d}_{\mathrm{b}}^{\infty}(\mathrm{PD}(f), \mathrm{PD}(R_{\delta\to\delta'}(P_\alpha, d_P))) \leq c\delta' + c\varepsilon_m(C) + \varepsilon_f(C)$$

Proof of Lemma 3.2.

Let $C \in \mathcal{C}(P, \mathbb{X})$ such that $\varepsilon_m(C) < \frac{1}{10}\varrho(\mathbb{X})$. Let $\pi : P \to \mathbb{X}$, such that

$$\forall p \in P, (p, \pi(p)) \in C$$

Let $L = \pi(P) \subseteq \mathbb{X}$. We assume, without loss of generality, that π is injective (we will see later on why this assumption can be made). Thus, it is a bijection between P and L. Thus, we can define a new distance $\tilde{d}_{\mathbb{X}}$ on $\mathbb{X}$:

$$\forall x, x' \in \mathbb{X}, \tilde{d}_{\mathbb{X}}(x, x') = d_P(\pi^{-1}(x), \pi^{-1}(x'))$$

$$
\begin{tikzcd}
 & R_{\delta'}(\tilde{L}_\alpha, \tilde{d}_X) \arrow[rr] & & R_{\delta'}(\tilde{L}_\beta, \tilde{d}_X) \\
R_\delta(\tilde{L}_\alpha, \tilde{d}_X) \arrow[ur] \arrow[rr] & & R_\delta(\tilde{L}_\beta, \tilde{d}_X) \arrow[ur] & \\
 & R_{\delta'}(P_\alpha, d_P) \arrow[uu] \arrow[rr] & & R_{\delta'}(P_\beta, d_P) \arrow[uu] \\
R_\delta(P_\alpha, d_P) \arrow[uu] \arrow[ur] \arrow[rr] & & R_\delta(P_\beta, d_P) \arrow[uu] \arrow[ur] &
\end{tikzcd}
$$

Figure 1: The commutative diagram formed by canonical inclusions (horizontal arrows) and induced simplicial isomorphisms (vertical arrows), for all values $\alpha \leq \beta \in \mathbb{R}$.

We have, by definition:

$$\forall x, x' \in \mathbb{X}, |d_{\mathbb{X}}(x, x') - \tilde{d}_{\mathbb{X}}(x, x')| \leq \varepsilon_m(C)$$

We claim L is an $\varepsilon_m(C)$-sample of $\mathbb{X}$. Indeed, let $x \in \mathbb{X}$ and $p \in P$ such that $(p, x) \in C$. Then

$$|d_{\mathbb{X}}(x, \pi(p)) - d_P(p, p)| = |d_{\mathbb{X}}(x, \pi(p))| \leq \varepsilon_m(C)$$

Let $\tilde{g} : L \to \mathbb{R}$ defined by $\tilde{g} = g \circ \pi^{-1}$. Let $\zeta = \max_{q \in L} \ |\tilde{g}(q) - f(q)|$. We also define $\tilde{L}_\alpha = \tilde{g}^{-1}((-\infty, \alpha])$ and the corresponding Vietoris-Rips nested persistence module $R_{\delta \to \delta'}(\tilde{L}_\alpha, d_{\mathbb{X}})$

Let us pick some δ, δ' as in the statement of the Lemma. Then, by the triangle inequality:

$$\begin{aligned} \mathrm{d}_{\mathrm{b}}^{\infty}(\mathrm{PD}(f), \mathrm{PD}(R_{\delta \to \delta'}(P_\alpha, d_P))) &\leq \mathrm{d}_{\mathrm{b}}^{\infty}(\mathrm{PD}(f), \mathrm{PD}(R_{\delta \to \delta'}(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}}))) \\ &+ \mathrm{d}_{\mathrm{b}}^{\infty}(\mathrm{PD}(R_{\delta \to \delta'}(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}})), \mathrm{PD}(R_{\delta \to \delta'}(P_\alpha, d_P))) \end{aligned}$$

We will now bound the two terms in the sum independently.

Second term.

Recall that we assumed the map $\pi : P \to L$ to be bijective. By definition of $\tilde{g} = g \circ \pi^{-1}$, for any $\alpha \in \mathbb{R}$ the restriction $\pi|_{g^{-1}((-\infty, \alpha])}$ is an isometry onto $\tilde{g}^{-1}((-\infty, \alpha])$, and so the induced simplicial maps $R_\delta(P_\alpha, d_P) \to R_\delta(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}})$ and $R_{\delta'}(P_\alpha, d_P) \to R_{\delta'}(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}})$ are isomorphisms. Moreover, these isomorphisms make the diagram of Figure 1 commute for all $\alpha \leq \beta \in \mathbb{R}$, so the persistence modules $R_{\delta \to \delta'}(P_\alpha, d_P)$ and $R_{\delta \to \delta'}(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}})$ are isomorphic. Their bottleneck distance is thus 0.

First term.

We want to use Lemma 3.1. Thus, we need to interleave the filtration $R_\delta(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}})$ between three Čech filtrations $C_{\varepsilon_m(C)}(\tilde{L}_\alpha, d_{\mathbb{X}})$, $C_{\varepsilon'}(\tilde{L}_\alpha, d_{\mathbb{X}})$ and $C_{\varepsilon''}(\tilde{L}_\alpha, d_{\mathbb{X}})$ such that $\varepsilon_m(C) < \varepsilon' < \varepsilon'' < \varrho(\mathbb{X})$. We use Equation (1) and

$$d_{\mathbb{X}}(x, x') - \varepsilon_m(C) \leq \tilde{d}_{\mathbb{X}}(x, x') \leq d_{\mathbb{X}}(x, x') + \varepsilon_m(C)$$

to get the following sequence of inclusions:

$$C_{\varepsilon_m(C)}(\tilde{L}_\alpha, d_{\mathbb{X}}) \subseteq R_{2\varepsilon_m(C)}(\tilde{L}_\alpha, d_{\mathbb{X}}) \subseteq R_{3\varepsilon_m(C)}(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}}) \overset{(i)}{\subseteq} R_\delta(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}})$$

$$\subseteq R_{\delta+\varepsilon_m(C)}(\tilde{L}_\alpha, d_{\mathbb{X}}) \subseteq C_{\delta+\varepsilon_m(C)=\varepsilon'}(\tilde{L}_\alpha, d_{\mathbb{X}}) \subseteq R_{2\varepsilon'}(\tilde{L}_\alpha, d_{\mathbb{X}})$$

$$\subseteq R_{2\varepsilon'+\varepsilon_m(C)}(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}}) \overset{(ii)}{\subseteq} R_{\delta'}(\tilde{L}_\alpha, \tilde{d}_{\mathbb{X}}) \subseteq R_{\delta'+\varepsilon_m(C)}(\tilde{L}_\alpha, d_{\mathbb{X}}) \overset{(iii)}{\subseteq} C_{\delta'+\varepsilon_m(C)=\varepsilon''}(\tilde{L}_\alpha, d_{\mathbb{X}})$$

We must have $\varepsilon' \leq \varepsilon'' \leq \varrho(\mathbb{X})$ to use Lemma 3.1.

– (i) imposes $3\varepsilon_m(C) \leq \delta$

– (ii) imposes $2\varepsilon' + \varepsilon_m(C) = 2\delta + 3\varepsilon_m(C) \leq \delta'$

– (iii) imposes $\delta' + \varepsilon_m(C) \leq \varrho(\mathbb{X})$

Finding a δ' such that inequalities (ii) and (iii) are verified is possible only if

$$(iv) \quad 2\delta + 3\varepsilon_m(C) \leq \varrho(\mathbb{X}) - \varepsilon_m(C) \Leftrightarrow \delta \leq \varrho(\mathbb{X})/2 - 2\varepsilon_m(C)$$

Finding a δ such that inequalities (i) and (iv) are verified is possible only if

$$3\varepsilon_m(C) \leq \varrho(\mathbb{X})/2 - 2\varepsilon_m(C) \Leftrightarrow \varepsilon_m(C) \leq \varrho(\mathbb{X})/10$$

These inequalities explain the lower and upper bounds for the parameters δ and δ' in the statement of the theorem, as well as the assumption on the Gromov-Hausdorff distance between P and $\mathbb{X}$.

Then, using Lemma 3.1, we can bound the term by $c\varepsilon'' + \zeta = c(\delta' + \varepsilon_m(C)) + \zeta$. Finally, as $\zeta \leq \varepsilon_f(C)$, the result follows.

To complete the proof, we now explain why the map $\pi : P \to \mathbb{X}$ can be assumed to be injective without loss of generality. Since the length space $\mathbb{X}$ has a finite inner metric $d_{\mathbb{X}}$, every pair of points is connected by a rectifiable path in $\mathbb{X}$. It follows that every open metric ball in $\mathbb{X}$ is infinite, provided that $\mathbb{X}$ itself is not reduced to a point. Leaving aside the special case where $\mathbb{X}$ is reduced to a point as an easy exercise, we conclude that for any $\eta > 0$ the points of $\pi(P)$ can be perturbed at will within distance η in $\mathbb{X}$ so as to make π injective. These perturbations may raise the metric and functional distortions of π, but no higher than $\varepsilon_m(C) + \zeta$ and $\varepsilon_f(C) + c\eta$ respectively, by the triangle inequality and by the c-Lipschitz continuity of f. Thus one would have to add $(c+1)\eta$ to the bound. However, as this can be done for arbitrarily small η, the result follows from the limit case $\eta \to 0$. $\square$

For instance, we can take $\delta' = 3\delta$ to immediately get the following corollary.

Lemma 3.3 *Let $(\mathbb{X}, d_{\mathbb{X}})$, (P, d_P), $f : \mathbb{X} \to \mathbb{R}$ and $g : P \to \mathbb{R}$ be as in Lemma 3.2. Then, for any correspondence $C \in \mathcal{C}(P, \mathbb{X})$ such that $\varepsilon_m(C) < \frac{\varrho(\mathbb{X})}{10}$, and for any parameter $\delta \in (3\varepsilon_m(C), \frac{1}{3}(\varrho(\mathbb{X}) - \varepsilon_m(C)))$,*

$$\mathrm{d}_\mathrm{b}^\infty(\mathrm{PD}(f), \mathrm{PD}(R_{\delta\to 3\delta}(P_\alpha, d_P))) \leq 3c\delta + c\varepsilon_m(C) + \varepsilon_f(C)$$

We can finally state the main theorem.

Theorem 3.4 *Let $\mathbb{X}$ and $\mathbb{Y}$ be two compact length spaces with curvature bounded above. Let $\varrho(\mathbb{X})$ and $\varrho(\mathbb{Y})$ be their convexity radii. Let $f : \mathbb{X} \to \mathbb{R}$ and $g : \mathbb{Y} \to \mathbb{R}$ be two Lipschitz functions with constants c_f and c_g respectively. Assume $\mathrm{d_{GH}}(\mathbb{X},\mathbb{Y}) \leq \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{20}$. Then, for any correspondence $C \in \mathcal{C}(\mathbb{X},\mathbb{Y})$ such that $\varepsilon_m(C) < \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{10}$*

$$\mathrm{d_b^\infty}(\mathrm{PD}(f),\mathrm{PD}(g)) \leq (9(c_f+c_g)+\min(c_f,c_g))\varepsilon_m(C)+\varepsilon_f(C) \tag{3}$$

Proof of Theorem 3.4. Let C be a correspondence between $\mathbb{X}$ and $\mathbb{Y}$ such that $\varepsilon_m(C) < \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{10}$. Let $0 < \mu < \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{20} - \frac{\varepsilon_m(C)}{2}$ and take a finite μ-sample P of $\mathbb{X}$. Let $P_\alpha = P \cap f^{-1}((-\infty,\alpha])$. We will apply Lemma 3.3 to both pairs $(P,\mathbb{X})$ and $(P,\mathbb{Y})$.

Firstly, we check the assumptions of the Lemma:

$$\mathrm{d_{GH}}(P,\mathbb{X}) \leq \mu \leq \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{20} \leq \frac{\varrho(\mathbb{X})}{20}$$

$$\mathrm{d_{GH}}(P,\mathbb{Y}) \leq \mu+\mathrm{d_{GH}}(\mathbb{X},\mathbb{Y}) \leq \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{20}+\left(\mathrm{d_{GH}}(\mathbb{X},\mathbb{Y})-\frac{\varepsilon_m(C)}{2}\right) \leq \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{20} \leq \frac{\varrho(\mathbb{Y})}{20}$$

Secondly, we now build two correspondences $C_\mathbb{X} \subseteq P \times \mathbb{X}$ and $C_\mathbb{Y} \subseteq P \times \mathbb{Y}$. Let

$$C_\mathbb{X} = \{(p,x) \mid p = \min_{q \in P} d_\mathbb{X}(q,x)\}$$

$$C_\mathbb{Y} = \{(p,y) \mid \exists x \in \mathbb{X} \text{ s.t. } (x,y) \in C \text{ and } (p,x) \in C_\mathbb{X}\}$$

Clearly, we have the following inequalities:
- $\varepsilon_m(C_\mathbb{X}) \leq 2\mu \leq \frac{\varrho(\mathbb{X})}{10}$
- $\varepsilon_m(C_\mathbb{Y}) \leq \varepsilon_m(C)+2\mu \leq \frac{\varrho(\mathbb{Y})}{10}$
- $\varepsilon_f(C_\mathbb{X}) \leq c_f\mu$
- $\varepsilon_f(C_\mathbb{Y}) \leq \varepsilon_f(C)+c_f\mu$

Finally, we pick some arbitrary $\delta \in (6\mu+3\varepsilon_m(C), \frac{1}{3}(\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))-2\mu-\varepsilon_m(C)))$ and we apply Lemma 3.3 to obtain:

$$\mathrm{d_b^\infty}(\mathrm{PD}(f),\mathrm{PD}(g)) \leq 3c_f\delta + c_f\varepsilon_m(C_\mathbb{X}) + \varepsilon_f(C_\mathbb{X}) + 3c_g\delta + c_g\varepsilon_m(C_\mathbb{Y}) + \varepsilon_f(C_\mathbb{Y})$$

$$= 3(c_f+c_g)\delta + c_g\varepsilon_m(C) + \varepsilon_f(C) + (4c_f+2c_g)\mu$$

By taking the limit case $\mu \to 0$ and $\delta \to 6\mu+3\varepsilon_m(C)$, we end up with the new bound

$$\mathrm{d_b^\infty}(\mathrm{PD}(f),\mathrm{PD}(g)) \leq (9(c_f+c_g)+c_g)\varepsilon_m(C)+\varepsilon_f(C)$$

As the problem is symmetric in $\mathbb{X}$ and $\mathbb{Y}$, the final result follows. □

Remark 1 *Let us consider two special cases.*
- *Assume $c_f = c_g = c$. Then we have*

$$\mathrm{d_b^\infty}(\mathrm{PD}(f),\mathrm{PD}(g)) \leq 19c\varepsilon_m(C)+\varepsilon_f(C)$$

– *Let $a \in \mathbb{X}$ and $b \in \mathbb{Y}$ be two source points. Assume $f(\cdot) = d_{\mathbb{X}}(a, \cdot)$ and $g(\cdot) = d_{\mathbb{Y}}(b, \cdot)$ with $(a, b) \in C$. Then $c_g = c_f = 1$ because f and g are 1-Lipschitz. Furthermore,*

$$\varepsilon_f(C) = \sup_{(x,y)\in C} |d_{\mathbb{X}}(a, x) - d_{\mathbb{Y}}(b, y)| \leq \varepsilon_m(C)$$

$$\Rightarrow \mathrm{d}_\mathrm{b}^\infty(\mathrm{PD}(f), \mathrm{PD}(g)) \leq 20\varepsilon_m(C)$$

Recall that the length spaces in this paper are assumed to have a finite length metric. It follows that the quantities $\varepsilon_m(C)$ and $\varepsilon_f(C)$ in the conclusion of the theorem are always finite. Thus, letting $f = 0$ and $g = 0$, we have $\mathrm{d}_\mathrm{b}^\infty(\mathrm{PD}(f),\ \mathrm{PD}(g) < +\infty$, which implies that the number of essential classes in both persistence diagrams is the same, or equivalently, that the homology groups of $\mathbb{X}$ and $\mathbb{Y}$ are isomorphic.

Corollary 3.5 (Homological stability) *Given two compact length spaces $(\mathbb{X}, d_{\mathbb{X}})$ and $(\mathbb{Y}, d_{\mathbb{Y}})$ of curvature bounded above, if $\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y}) \leq \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{20}$ then $\forall k \in \mathbb{N}$, $H_k(\mathbb{X}) \simeq H_k(\mathbb{Y})$.*

Note that the condition that the length spaces $\mathbb{X}$ and $\mathbb{Y}$ have large convexity radii compared to $\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ is important for the conclusions of the theorem and corollary to hold. It is indeed easy to build counter-examples in which one of the spaces has a small convexity radius and the conclusions fail to hold. Take for instance for $\mathbb{X}$ the unit circle in the plane, and for $\mathbb{Y}$ the unit open segment, both equipped with the intrinsic metrics induced by the Euclidean distance in $\mathbb{R}^2$. Then, although their Gromov-Hausdorff distance is finite and the convexity radius of $\mathbb{Y}$ is infinite, their 1-dimensional homology groups are not isomorphic, and as a result for any Lipschitz functions $f : \mathbb{X} \to \mathbb{R}$ and $g : \mathbb{Y} \to \mathbb{R}$ the bottleneck distance $\mathrm{d}_\mathrm{b}^\infty(\mathrm{PD}(f), \mathrm{PD}(g))$ is infinite, the numbers of essential 1-dimensional homology classes in both diagrams being different.

In practice, most of the time the inputs come from finite metric spaces. One can derive an analogous version of Theorem 3.4 for them.

Theorem 3.6 *Let $\mathbb{X}$ and $\mathbb{Y}$ be two compact length spaces with curvature bounded above. Let $\varrho(\mathbb{X})$ and $\varrho(\mathbb{Y})$ be their convexity radii. Let $f : \mathbb{X} \to \mathbb{R}$ and $g : \mathbb{Y} \to \mathbb{R}$ be two c Lipschitz functions. Assume $\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y}) \leq \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{20}$. Let $C \in \mathcal{C}(\mathbb{X}, \mathbb{Y})$ s.t. $\varepsilon_m(C) < \frac{\min(\varrho(\mathbb{X}),\varrho(\mathbb{Y}))}{10}$. Let*

– $\mu \in (0, \frac{\varrho(\mathbb{X})}{20} - \frac{\varepsilon_m(C)}{2}]$

– $\delta \in (3\varepsilon_m(C) + 6\mu, \frac{1}{3}\min(\varrho(\mathbb{X}), \varrho(\mathbb{Y})) - \varepsilon_m(C) - 2\mu)$

Then, for any finite μ-sample P of $\mathbb{X}$ and μ-sample Q of $\mathbb{Y}$

$$\mathrm{d}_\mathrm{b}^\infty(\mathrm{PD}(R_{\delta\to3\delta}(P_\alpha, d_{\mathbb{X}})), \mathrm{PD}(R_{\delta\to3\delta}(Q_\alpha, d_{\mathbb{Y}}))) \leq 19c\varepsilon_m(C) + \varepsilon_f(C) + 42c\mu \tag{4}$$

where $P_\alpha = P \cap f^{-1}((-\infty, \alpha])$ and $Q_\alpha = Q \cap g^{-1}((-\infty, \alpha])$.

Proof of Theorem 3.6. Apply Lemma 3.3 to pairs $(P, \mathbb{X})$ and $(Q, \mathbb{X})$. □

Thus, we have seen how one can derive global persistence diagrams (PDs) from general Lipschitz functions, and local PDs from Lipschitz functions anchored at source points. In the next section, we show how to turn these PDs into stable signatures in $\mathbb{R}^d$.

3.2 Topological Signatures

Mapping PDs to Euclidean spaces can be of great interest in many applications, as the space of PDs is not suited for the computation of basic quantites such as mean or variance.

Definition 3.7 *Let* PD *be an arbitrary finite persistence diagram, and let*

$$S = \{\min(\|p-q\|_\infty, p_\Delta(p), p_\Delta(q)) \mid p, q \in \mathrm{PD}\}$$

where $p_\Delta(\cdot)$ denotes the distance (with the infinity norm) to the diagonal. The topological signature *$V \in \mathbb{R}^{|S|}$ is the vector of the elements of S sorted with decreasing order. If there is only one point in* PD*, we arbitrary set $V = 0$.*

Theorem 3.8 *Let PD_x and PD_y be two finite persistence diagrams and V_x and V_y be their associated topological signatures. Let $N_x = |\mathrm{PD}_x|$, $N_y = |\mathrm{PD}_y|$ and $N = \max(N_x, N_y)$. Then*

$$\sqrt{\frac{2}{N(N-1)}}\|V_x - V_y\|_2 \leq \|V_x - V_y\|_\infty \leq 2\mathrm{d}_\mathrm{b}^\infty(\mathrm{PD}_x, \mathrm{PD}_y)$$

Proof of Theorem 3.8. Let $\varepsilon = \mathrm{d}_\mathrm{b}^\infty(\mathrm{PD}_x, \mathrm{PD}_y)$. As the problem is symmetric in x and y, assume without loss of generality that $N_x < N_y$. We consider one of the matching γ^* realizing the bottleneck distance between PD_x and PD_y. We also call $N_{x,1}$ (resp. $N_{x,2}$) the number of points of PD_x which are mapped by γ^* to an element of PD_y (resp. to the diagonal). We have $N_{x,1} + N_{x,2} = N_x$. Thus, $N_{x,1}$ elements of PD_y are mapped to elements of PD_x, $N_{x,2}$ points to the diagonal, and the $N_y - N_x$ other elements of PD_y are also mapped to the diagonal. We introduce a bijective mapping $\psi : \mathrm{PD}_x \to \mathbb{R}^2$ which coincides with γ^* on the $N_{x,1}$ points of PD_x which are not mapped to the diagonal and which arbitrarily associates the remaining $N_{x,2}$ elements of PD_x to the corresponding $N_{x,2}$ points of PD_y.

By definition, we have

$$V_x = [\ \min\ \{\|p_i - p_j\|_\infty, p_\Delta(p_i), p_\Delta(p_j)\}\ ]_{1 \leq i,j \leq N_x}$$

and $(V_x)_i \geq (V_x)_{i+1},\ \forall i \in [1, N_x(N_x-1)/2 - 1]$.

Let $\hat{V}_y$ be

$$\hat{V}_y = [\ \min\ \{\|\psi(p_i) - \psi(p_j)\|_\infty, p_\Delta(\psi(p_i)), p_\Delta(\psi(p_j))\}\ ]_{1 \leq i,j \leq N_x}$$

Then, we add the remaining pairwise terms of PD_y in $\hat{V}_y$ and we also fill V_x with null values until its length is $N_y(N_y-1)/2$ so that both vectors have the same size.

We will now prove $\|V_x - \hat{V}_y\|_\infty \leq 2\varepsilon$.

Namely, we have $(V_x)_i = \min\ \{\|x_{i,1} - x_{i,2}\|_\infty, p_\Delta(x_{i,1}), p_\Delta(x_{i,2})\}$ or 0, and $(\hat{V}_y)_i = \min\ \{\|y_{i,1} - y_{i,2}\|_\infty, p_\Delta(y_{i,1}), p_\Delta(y_{i,2})\}$. We have three different cases to treat here:

- (a) $i \leq \frac{N_x(N_x-1)}{2}$ and the two pairs of points are matched by the bottleneck matching
- (b) $i \leq \frac{N_x(N_x-1)}{2}$ and at least one point of each pair is matched to the diagonal
- (c) $i > \frac{N_x(N_x-1)}{2}$, then $(V_x)_i = 0$

Case (c) is easy to treat. Indeed, we know that at least one of the points of the pairwise term in $(\hat{V}_y)_i$, say $y_{i,1}$, is matched to the diagonal. Thus, we have

$$|(V_x)_i - (\hat{V}_y)_i| = |(\hat{V}_y)_i| \leq |p_\Delta(y_{i,1})| \leq \varepsilon \leq 2\varepsilon$$

Case (b). Here we know that at least one point of the pairwise term in $(V_x)_i$, say $x_{i,1}$, and one of the pairwise term in $(\hat{V}_y)_i$, say $y_{i,1}$, are mapped to the diagonal, the other being either mapped together or also to the diagonal. Then

$$|(V_x)_i - (\hat{V}_y)_i| \leq |p_\Delta(x_{i,1})| + |p_\Delta(y_{i,1})| \leq 2\varepsilon$$

Case (a). Here we have $\gamma^*(x_{i,1}) = y_{i,1}$ and $\gamma^*(x_{i,2}) = y_{i,2}$. Three different sub cases come out:

- (i) The minimum is in both cases the distance between the points. Then we have $|(V_x)_i - (\hat{V}_y)_i| = |\|x_{i,1} - x_{i,2}\|_\infty - \|y_{i,1} - y_{i,2}\|_\infty| \leq 2\varepsilon$
- (ii) The minimum is in both cases the distance of a point to the diagonal. Then either $|(V_x)_i - (\hat{V}_y)_i| = |p_\Delta(x_{i,1}) - p_\Delta(y_{i,1})|$ or $|(V_x)_i - (\hat{V}_y)_i| = |p_\Delta(x_{i,1}) - p_\Delta(y_{i,2})|$.
  The first case is easy, and the bound is immediate as the points are mapped by γ^*.
  Second case is trickier. We have the following inequalities:
  - $\eta = p_\Delta(x_{i,2}) - p_\Delta(x_{i,1}) \geq 0$
  - $p_\Delta(y_{i,1}) = p_\Delta(x_{i,1}) + \alpha_1$ with $|\alpha_1| \leq \varepsilon$
  - $p_\Delta(y_{i,2}) = p_\Delta(x_{i,2}) + \alpha_2$ with $|\alpha_2| \leq \varepsilon$

  Thus $\varepsilon \geq \alpha_1 \geq \alpha_2 + \eta \geq \eta - \varepsilon \geq -\varepsilon$ and

  $$|(V_x)_i - (\hat{V}_y)_i| = |p_\Delta(x_{i,1}) - p_\Delta(y_{i,2})| = |\eta + \alpha_2| \leq \varepsilon \leq 2\varepsilon$$

- (iii) The minimum is the distance of a point to the diagonal for one term and the distance between the points for the other, say

  $$d_x = \|x_{i,1} - x_{i,2}\|_\infty \leq p_\Delta(x_{i,1}), p_\Delta(x_{i,2})$$

  $$p_\Delta(y_{i,1}) \leq d_y = \|y_{i,1} - y_{i,2}\|_\infty, p_\Delta(y_{i,2})$$

  Then $|(V_x)_i - (\hat{V}_y)_i| = |d_x - p_\Delta(y_{i,1})|$. As $p_\Delta(y_{i,1}) \geq p_\Delta(x_{i,1}) - \varepsilon$, we have

  $$d_x - p_\Delta(y_{i,1}) \leq \varepsilon + (d_x - p_\Delta(x_{i,1})) \leq \varepsilon \leq 2\varepsilon$$

  We also have

  $$p_\Delta(y_{i,1}) \leq d_y \leq d_x + 2\varepsilon$$

  And thus

  $$|(V_x)_i - (\hat{V}_y)_i| \leq 2\varepsilon$$

Thus, $\|V_x - \hat{V}_y\|_\infty \leq 2\varepsilon$. Now we prove and use the following lemma to conclude:

Lemma 3.9 *Let $U, V \in \mathbb{R}_+^n$. We suppose that U is decreasing (i.e. $\forall i \in \{1 \dots n-1\}$, we have $U_i \geq U_{i+1}$) and that $\|U - V\|_\infty \leq \alpha$. $\tilde{V} \in \mathbb{R}_+^n$ is the image of V by a coordinate permutation σ which makes him decreasing (i.e. $\forall i \in \{1 \dots n-1\}$, we have $\tilde{V}_i = V_{\sigma(i)}$ and $\tilde{V}_i \geq \tilde{V}_{i+1}$). Then*

$$\|U - \tilde{V}\|_\infty \leq \alpha$$

Proof of Lemma 3.9. Because of $\|U - V\|_\infty \le \alpha$, we can always bound an element v_i of V between $u_i - \alpha$ and $u_i + \alpha$.

Let $i \in \{1 \dots n\}$ and $v_i = u_i + x_i$ where $-\alpha \le x_i \le \alpha$. When we take the infinite norm of $U - \tilde{V}$, we have to estimate v_i's position in $\tilde{V}$. If we define

$$j_i = \min \{t > i \mid u_t + \alpha < v_i\}$$

(or $j_i = n + 1$ if the set is empty) and

$$k_i = \max \{t < i \mid u_t - \alpha > v_i\}$$

(or $k_i = 0$ if the set is empty) then we know that v_i's position in $\tilde{V}$ is a unique integer l_i between $k_i + 1$ and $j_i - 1$.

There are two cases to consider: either $i \ge l_i \ge j_i - 1$, and we have $u_{l_i} + \alpha \ge u_i + x_i$, thus

$$|u_i - u_{l_i} + x_i| = |v_i - u_{l_i}| \le \alpha$$

If $i \le l_i \le k_i + 1$, the proof is exactly the same.

This inequality being true for every i, it is also true for the vectors in the infinite norm and the proof is over. $\square$

We can then finally conclude : $\|V_x - V_y\|_\infty \le 2\varepsilon$ $\square$

References

[1] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[2] M. Carrière, S. Y. Oudot, and M. Ovsjanikov. Stable topological signatures for points on 3d shapes. *Proc. SGP*, 2015. To appear.

[3] F. Chazal, V. de Silva, M. Glisse, and S. Y. Oudot. The structure and stability of persistence modules. *CoRR*, abs/1207.3674, 2012.

[4] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba. Analysis of scalar fields over point cloud data. Research Report 6576, INRIA, July 2008.

[5] F. Chazal and S. Y. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. Comput. Geom.*, pages 232–241, 2008.

[6] A. Hatcher. *Algebraic Topology.* Cambridge Univ. Press, 2001.

[7] J. R. Munkres. *Elements of algebraic topology.* Addison-Wesley, Redwood City, California, 1984.