Lipschitz Functions Have L_p-stable Persistence *

David Cohen-Steiner†, Herbert Edelsbrunner‡, John Harer§ and Yuriy Mileyko¶

Abstract

We prove two stability results for Lipschitz functions on triangulable, compact metric spaces and consider applications of both to problems in systems biology. Given two functions, the first is formulated in terms of the Wasserstein distance between their persistence diagrams and the second in terms of their total persistences.

Keywords. Continuous functions, comparison, classification, metric spaces, persistent homology, Wasserstein distance, total persistence, stability, gene expression.

1 Introduction

The mathematical methods developed in this paper are motivated by biological questions of gene regulation. Measuring gene expression over time, we get functions and we are interested in quantifying their shape and similarity. Returning to the biological questions in Section 4, we now focus on the mathematical results which are vastly more general than what we need for the specific applications. Specifically, let $\mathbb{X}$ be a triangulable, compact metric space and $f, g : \mathbb{X} \to \mathbb{R}$ two tame Lipschitz functions. Then there exist constants k and C that depend on $\mathbb{X}$ and on the Lipschitz constants of f and g such that the degree-p Wasserstein distance between the corresponding persistence diagrams is

$$W_p(f,g) \quad \leq \quad C \cdot \|f-g\|_\infty^{1-\frac{k}{p}}$$

for every $p \geq k$. The second result pertains to the sum of p-th powers of persistences, referred to as the degree-p total persistence and denoted as $\mathrm{Pers}_p(f)$. We prove there are constants k and C such that

$$|\mathrm{Pers}_p(f) - \mathrm{Pers}_p(g)| \quad \leq \quad C \cdot \|f-g\|_\infty$$

for every $p \geq k+1$. More complete versions of both inequalities are stated as the Wasserstein and the Total Persistence Stability Theorems in Section 3.

To put the two results in perspective, we recall recent work on persistent homology. Letting $\mathbb{X}_a = f^{-1}(-\infty, a]$ be the sublevel set defined by the threshold a, we consider the homology groups of $\mathbb{X}_a$, and for $a \leq b$ we consider the homomorphisms between the corresponding homology groups induced by the inclusion $\mathbb{X}_a \subseteq \mathbb{X}_b$. As shown in [7], the homology classes can be tracked within this sequence of homology groups and it is possible to determine the moments of birth and death for each; see also [9, 15]. The corresponding persistence diagram is a multi-set of points in two dimensions in which each point represents a homology class, marking its birth by the first and its death but the second coordinate. A breakthrough result is the stability of these diagrams proved in [3]. Specifically, the bottleneck distance, defined as the maximum L_∞-distance between two matched points of the diagrams of f and of g, is at most $\|f-g\|_\infty$. Of course it is possible that the diagrams of f and of g have different size, but we can always use points on the diagonal to complete or improve the matching.

The two results in this paper go beyond this stability theorem by measuring similarity in terms of sums of powers. This provides the sensitivity to local variation. Both inequalities are false for general continuous functions and crucially rely on the assumption that f and g are Lipschitz.

Outline. Section 2 looks at triangulations of $\mathbb{X}$ with small mesh and small size and uses them to prove an upper bound on the number of homology classes of persistence beyond some threshold. Section 3 establishes the two stability results. Section 4 sketches the applications of the inequalities to analyzing gene expression in development. Section 5 concludes the paper.

*This research is partially supported by the Defense Advanced Research Projects Agency (DARPA) under grants HR0011-05-1-0007 and HR0011-05-1-0057 and by CNRS under grant PICS-3416.

†INRIA, 2004 Route des Lucioles, BP93, Sophia-Antipolis, France.

‡Departments of Computer Science and of Mathematics, Duke University, Durham, and Geomagic, Research Triangle Park, North Carolina, USA.

§Department of Mathematics and Section in Computational Biology and Bioinformatics, Duke University, Durham, North Carolina, USA.

¶Department of Computer Science, Duke University, Durham, North Carolina, USA.

2 Preliminaries

In this section, we introduce the tools we need to prove our results, triangulations of small mesh and combinatorial arguments that relate the number of homology classes to the sizes of these triangulations. We assume basic familiarity with homology theory as described in [12].

2.1 Triangulations

We begin by introducing the type of triangulation we need and by explaining to what extent cycles in $\mathbb{X}$ can be replaced by cycles in the appropriate skeleton of the triangulation.

Mesh and size. Let $\mathbb{X}$ be a triangulable, compact metric space and $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ its metric. A *triangulation* of $\mathbb{X}$ is a finite simplicial complex K together with a homeomorphism from the underlying space of the complex to the space, $\vartheta : |K| \to \mathbb{X}$. We define the *diameter* of a simplex σ in K as the maximum distance between any two points of its image, $\operatorname{diam}(\sigma) = \max_{x,y\in\sigma} d(\vartheta(x), \vartheta(y))$. The *mesh* of the triangulation is the largest diameter, $\operatorname{mesh}(K) = \max_{\sigma\in K} \operatorname{diam}(\sigma)$. The *size* of the triangulation is the number of simplices, $\operatorname{card} K$. If the dimension of $\mathbb{X}$ is n then K consists of simplices of dimension between 0 and n. For each $0 \le \ell \le n$, the *ℓ-skeleton*, denoted as $K^{(\ell)}$, is the subset of simplices of dimension at most ℓ. Given a positive real number r, we are interested in a triangulation whose mesh is at most r and whose size is as small as possible. Specifically, we define

$$\begin{aligned} N(r) &= \min_{\operatorname{mesh}(K)\le r} \operatorname{card} K; \\ N_\ell(r) &= \min_{\operatorname{mesh}(K)\le r} (\operatorname{card} K^{(\ell)} - \operatorname{card} K^{(\ell-1)}), \end{aligned}$$

for each $0 \le \ell \le n$. Note that $\sum_{\ell=0}^{n} N_\ell(r) \le N(r)$.

Consider the n-dimensional sphere, $\mathbb{X} = \mathbb{S}^n$, as an example. For $n = 1$ its length is 2π and we can decompose it into $m = \lceil 2\pi/r \rceil$ arcs of length at most r each. Assuming $r < \pi$ we can triangulate with m edges and m vertices. It is impossible to do better, hence $N_0(r) = N_1(r) = m$ and $N(r) = 2m$. For $n \ge 2$ it is not as easy to pin down the functions but it is not difficult to see that there are positive constants c and C that depend on n and the n-dimensional volume of $\mathbb{S}^n$ such that $c/r^n \le N(r) \le C/r^n$. If $\mathbb{X}$ is a compact Riemannian n-manifold then N exhibits the same behavior for sufficiently small r but can be significantly different for large values of r. As an example, let $\mathbb{X}$ be the 2-dimensional torus consisting of all points at Euclidean distance $\varepsilon > 0$ from a circle of radius $\frac{1}{\varepsilon}$ in $\mathbb{R}^3$. Its area is $4\pi^2$. For small ε and relatively much larger r the number of simplices required to triangulate $\mathbb{X}$ is proportional to $\frac{1}{r}$, while for $r \le \varepsilon$ it is proportional to $\frac{1}{r^2}$.

Snapping. Let K be a triangulation of $\mathbb{X}$ with mesh equal to r. For a subset $z \subseteq \mathbb{X}$, we consider its thickened version, z^r, consisting of all points $x \in \mathbb{X}$ for which there is a point $y \in z$ with $d(x, y) \le r$. As illustrated in Figure 1, the thickened version of z contains a subset of the skeleton of the appropriate dimension that approximates z, both geometrically and homologically.

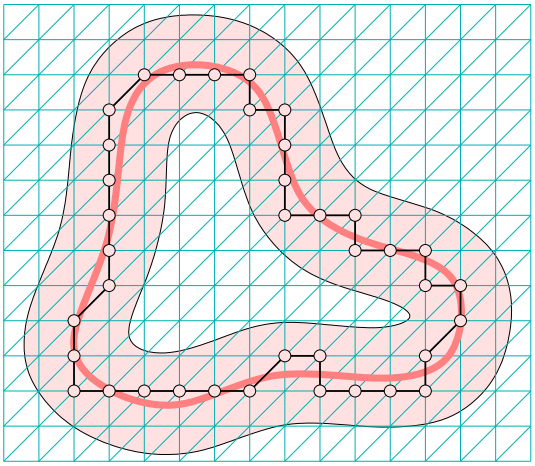

Figure 1: The 1-cycle in the edge-skeleton of the triangulation is homologous to the shaded 1-cycle inside the annulus around that 1-cycle.

SNAPPING LEMMA. Let K be a triangulation of a compact metric space $\mathbb{X}$ with $\operatorname{mesh}(K) = r$. Then for each cycle z of dimension ℓ in $\mathbb{X}$ there is a cycle $\bar{z}$ in the ℓ-skeleton of K that is homologous to z inside z^r.

PROOF. Let K_1 be the minimal subcomplex of K containing z. Since the map that sends each simplicial cycle of K_1 to its associated singular cycle induces an isomorphism at the homology level, we have that z is homologous to a simplicial cycle within K_1. Since all simplices in K_1 have diameter at most r, K_1 is included in z^r, which implies the result. □

Lipschitz functions. Let now $f : \mathbb{X} \to \mathbb{R}$ be a *Lipschitz function* on $\mathbb{X}$, that is, there is a constant c such that $f(x) - f(y) \le cd(x, y)$ for all $x, y \in \mathbb{X}$. The infimum over all such c is the *Lipschitz constant* of f, denoted as $\operatorname{Lip}(f)$. A crucial property of Lipschitz functions is that their level sets are well separated.

SEPARATION LEMMA. The distance between the level sets defined by values $a \le b$ is at least the difference in values divided by the Lipschitz constant, $d(x, y) \ge (b - a)/\operatorname{Lip}(f)$ whenever $f(x) = a$ and $f(y) = b$.

The proof is obvious. The Separation Lemma implies that the sublevel set $\mathbb{X}_b$ contains the thickened version $\mathbb{X}_a^r$ consisting of all points at distance at most $r = (b - a)/\operatorname{Lip}(f)$

from $\mathbb{X}_a$. The *amplitude* of f is the maximum difference in function values,

$$\mathrm{Amp}(f) \quad = \quad \max_{x\in\mathbb{X}} f(x) - \min_{y\in\mathbb{X}} f(y).$$

For a Lipschitz function, the amplitude is bounded from above by the diameter of the space times the Lipschitz constant, $\mathrm{Amp}(f) \leq \mathrm{diam}(\mathbb{X})\mathrm{Lip}(f)$.

2.2 Persistent Homology

We now bound the number of cycles whose persistence exceeds some threshold and we use this to bound sums of powers of persistences. These bounds will be instrumental in deriving our two main results in Section 3.

Birth and death. As before, we assume a compact metric space $\mathbb{X}$ and a Lipschitz function $f : \mathbb{X} \to \mathbb{R}$. For each value $a \in \mathbb{R}$, we have the sublevel set $\mathbb{X}_a = f^{-1}(-\infty, a]$ consisting of points with function value at most a. For $a \leq b$ we have $\mathbb{X}_a \subseteq \mathbb{X}_b$. This inclusion implies homomorphisms from the homology groups of $\mathbb{X}_a$ to those of $\mathbb{X}_b$,

$$\mathbf{f}_\ell^{a,b} : \mathsf{H}_\ell(\mathbb{X}_a) \to \mathsf{H}_\ell(\mathbb{X}_b),$$

one for each dimension ℓ. Throughout this paper, we assume our homology groups are defined for modulo-2 arithmetic but everything we say also holds for coefficient groups that are fields [15]. The nested family of sublevel sets defines a sequence of homology groups connected by the described homomorphisms. Following [3], we call f *tame* if this sequence is finite and consists of homology groups whose ranks are finite.

Within this framework, we can increase a from 0 to ∞ and observe homology classes appear and disappear. Specifically, a class $\alpha \in \mathsf{H}_\ell(\mathbb{X}_a)$ is *born* at $\mathbb{X}_a$ if α is not in the image of $\mathbf{f}_\ell^{a-\delta,a}$ for any $\delta > 0$. Furthermore, the class α born at $\mathbb{X}_a$ *dies entering* $\mathbb{X}_b$ if $\mathbf{f}_\ell^{a,b-\delta}(\alpha)$ is not in the image of $\mathbf{f}_\ell^{a-\delta,b-\delta}$, for any $\delta > 0$, but $\mathbf{f}_\ell^{a,b}(\alpha)$ is in the image of $\mathbf{f}_\ell^{a-\delta,b}$. If α is born at $\mathbb{X}_a$ and dies entering $\mathbb{X}_b$ then we set $\mathrm{b}(\alpha) = a$ and $\mathrm{d}(\alpha) = b$. It is also possible that α does not die in the sequence which ends with $\mathsf{H}_\ell(\mathbb{X})$. In this case, we set $\mathrm{d}(\alpha) = \max_{x\in\mathbb{X}} f(x)$. In summary, we have a value $\mathrm{b}(\alpha)$ and a value $\mathrm{d}(\alpha)$ for each cycle α that makes an appearance in the sequence of homology groups. The *persistence* of the class is the difference between the two values, $\mathrm{pers}(\alpha) = \mathrm{d}(\alpha) - \mathrm{b}(\alpha)$. This agrees with the original definition except for classes that do not die for which [7] set the persistence to infinity. The motivation for the slight change in definition is convenience. Without it, our theorems would require the more sophisticated concept of extended persistence as introduced in [4].

Persistence diagrams. We represent the births and deaths of ℓ-dimensional homology classes by a set of points in $\mathbb{R}^2$, the *ℓ-th persistence diagram* denoted as $\mathrm{Dgm}_\ell(f)$. For each class α that makes an appearance in the sequence of ℓ-th homology groups, the diagram contains the point $(\mathrm{b}(\alpha), \mathrm{d}(\alpha))$. We thus draw births along the horizontal axis, deaths along the vertical axis, and since deaths happen only after births, all points lie above the diagonal. In degenerate cases, classes can be born at the same time and die at the same time. Points in the diagram thus have integer multiplicities and $\mathrm{Dgm}_\ell(f)$ is a multiset.

This brings up an important subtlety about the meaning of a point in the diagram. With α an entire coset of homology classes is born at $a = \mathrm{b}(\alpha)$ and dies entering $b = \mathrm{d}(\alpha)$. Specifically, every class $\alpha + \beta$ with $\beta \in \mathsf{H}_\ell(\mathbb{X}_{a-\delta})$ is in this coset. The point (a, b) in $\mathrm{Dgm}_\ell(f)$ represents all these classes. It is a single point indicating that the rank of the ℓ-th homology group goes up by one at a and it drops by one at b. This is the case in which the rank changes by the appearance or disappearance of a single generator. In a degenerate case, we may have more than one generator appear at the same critical value. Say $a = \mathrm{b}(\alpha_1) = \mathrm{b}(\alpha_2) = \ldots = \mathrm{b}(\alpha_k)$ and all these classes are independent. Then the coset born at $\mathbb{X}_a$ consists of all classes $\alpha + \beta$, where α is a non-zero combination of the α_i and $\beta \in \mathsf{H}_\ell(\mathbb{X}_{a-\delta})$. The rank of the ℓ-th homology group increases by k. Correspondingly, the total multiplicity of the points with birth-coordinate a is k. The α_i may die at the same or at different critical values. Assume we have them indexed such that $\mathrm{d}(\alpha - 1) \leq \mathrm{d}(\alpha_2) \leq \ldots \leq \mathrm{d}(\alpha_k)$. For each non-zero combination $\alpha = \sum_{i=1}^{k} a_i\alpha_i$, we have a largest index $l = l(\alpha)$ such that $a_l = 1$. Then $\alpha + \beta$ dies at $\mathrm{d}(\alpha_l)$, the critical value at which its last constituent dies. Similar to births, we can have an arbitrary number, k, of independent classes die at the same value, b. Correspondingly, the rank of the ℓ-th homology group drops by k and the total multiplicity of points with death-coordinate b is k.

While there is nothing canonical about the choice of generators at births and deaths, the persistence diagram is independent of that choice and thus unique.

2.3 Persistent Cycles

We use the existence of triangulations with small mesh and small size to prove upper bounds on the sums of powers of persistences.

Number of cycles. We begin with bounding the number of homology classes of large persistence.

PERSISTENT CYCLE LEMMA. Let $\mathbb{X}$ be a triangulable, compact metric space and $f : \mathbb{X} \to \mathbb{R}$ a tame Lipschitz function. Then the number of points in the persistence diagrams of f whose persistence exceeds ε is at most $N(\varepsilon/\mathrm{Lip}(f))$.

PROOF. Set $r = \varepsilon/\mathrm{Lip}(f)$ and let K be a triangulation of $\mathbb{X}$ with $\mathrm{mesh}(K) \leq r$ and number of ℓ-simplices equal to $N_\ell(r)$. Let α be an ℓ-dimensional homology class whose persistence exceeds the threshold, $\mathrm{pers}(\alpha) = \mathrm{d}(\alpha) - \mathrm{b}(\alpha) >$

ε. By definition of birth, there is a cycle $z(\alpha)$ in $\mathbb{X}_{\mathrm{b}(\alpha)}$ that generates α in $\mathsf{H}_\ell(\mathbb{X}_{\mathrm{b}(\alpha)})$. Since f is Lipschitz, $\mathbb{X}_{\mathrm{b}(\alpha)+\varepsilon}$ contains $\mathbb{X}^r_{\mathrm{b}(\alpha)}$. The Snapping Lemma thus implies the existence of a cycle $\bar{z}(\alpha)$ in the ℓ-skeleton of K that is homologous to $z(\alpha)$ in $\mathbb{X}_{\mathrm{b}(\alpha)+\varepsilon}$.

Using this construction we obtain a cycle $\bar{z}_i = \bar{z}(\alpha_i)$ for each point in $\mathrm{Dgm}_\ell(f)$ whose persistence exceeds ε, α_i being an ℓ-dimensional homology class associated to that point. If points share the birth value then we chose the classes independent. Assuming the indices are chosen so that the birth values are ordered, we have $\mathrm{b}(\alpha_1) \leq \mathrm{b}(\alpha_2) \leq \ldots \leq \mathrm{b}(\alpha_m)$. A crucial property of this construction is the Independence of the m cycles in the ℓ-skeleton of K. We prove this by induction. Suppose that $\bar{z}_i$ is not independent of its predecessors. Then it is homologous to a linear combination, $\bar{z}_i \sim \sum_{j=1}^{i-1} a_j \bar{z}_j$. By construction, we have $\bar{z}_j \sim z(\alpha_j)$ in $\mathbb{X}_{\mathrm{b}(\alpha_i)+\varepsilon}$ for each $1 \leq j \leq i$. Thus,

$$z(\alpha_i) \quad \sim \quad \sum_{j=1}^{i-1} a_j z(\alpha_j)$$

in $\mathbb{X}_{\mathrm{b}(\alpha_i)+\varepsilon}$. If $\mathrm{b}(\alpha_{i-1}) < \mathrm{b}(\alpha_i)$ then this contradicts the assumption that the death of α_i comes strictly after $\mathrm{b}(\alpha_i) + \varepsilon$. Else let $k < i$ be the smallest index such that $\mathrm{b}(\alpha_k) = \mathrm{b}(\alpha_i)$. Then we have a non-zero combination of $\bar{z}_k$ to $\bar{z}_i$ homologous to $\sum_{k=1}^{k-1} a_j \bar{z}_j$. By construction, the death of the corresponding non-zero combination of the classes α_j is no earlier than the death of α_i. This again contradicts the assumption that $\mathrm{d}(\alpha_i) > \mathrm{b}(\alpha_i) + \varepsilon$.

Being independent, the number of cycles $\bar{z}_i$ is at most the rank of the ℓ-th homology group of the ℓ-skeleton, which is at most the number of ℓ-simplices, $m \leq \operatorname{rank} \mathsf{H}_\ell(K^{(\ell)}) \leq N_\ell(r)$. The claimed inequality follows because $N(r)$ is at least the sum of the $N_\ell(r)$ over all dimensions ℓ. □

Persistence moments. We are interested in the sum of k-th powers of the persistences of all points in the diagrams of f whose persistence exceeds some non-negative threshold,

$$\mathrm{Pers}_k(f,t) \quad = \quad \sum_{\mathrm{pers}(x)>t} \mathrm{pers}(x)^k.$$

For $t = 0$ we call $\mathrm{Pers}_k(f) = \mathrm{Pers}_k(f,0)$ the *degree-k total persistence* of f. We use the bound on the number of persistent cycles to get a bound on this sum.

MOMENT LEMMA. Let $\mathbb{X}$ be a triangulable, compact metric space and $f : \mathbb{X} \to \mathbb{R}$ a tame Lipschitz function. Then $\mathrm{Pers}_k(f,t)$ is bounded from above by

$$t^k N\left(\frac{t}{\mathrm{Lip}(f)}\right) + k \int_{\varepsilon=t}^{\mathrm{Amp}(f)} N\left(\frac{\varepsilon}{\mathrm{Lip}(f)}\right) \varepsilon^{k-1} \,\mathrm{d}\varepsilon.$$

PROOF. Let $P(\varepsilon)$ be the number of points in the diagrams of f whose persistence exceeds ε. By the Persistent Cycle Lemma we have $P(\varepsilon) \leq N(\varepsilon/\mathrm{Lip}(f))$. The derivative of P is the negative of the persistence values viewed as a distribution, that is, a sum of Dirac masses obtained by projecting the diagrams onto the anti-diagonal and scaling by a factor $\sqrt{2}$, as sketched in Figure 2. It follows that the sum of k-th

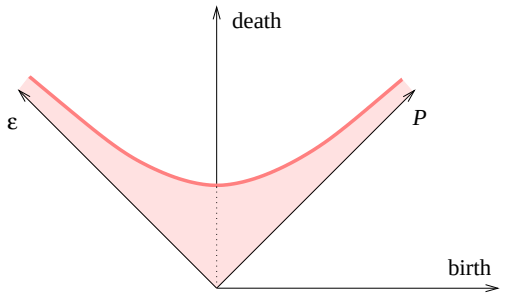

Figure 2: A cubistic sketch of the function P and its relation to the persistence diagrams. Imagine the P-axis normal to the birth,death-plane and the graph of the function drawn in the ε, P-plane.

powers of the persistences exceeding t is

$$\begin{aligned}\mathrm{Pers}_k(f,t) &= \int_{\varepsilon=t}^{\infty} -\frac{\partial P}{\partial \varepsilon}(\varepsilon)\varepsilon^k \,\mathrm{d}\varepsilon \\ &= [-P(\varepsilon)\varepsilon^k]_{\varepsilon=t}^{\infty} + \int_{\varepsilon=t}^{\infty} P(\varepsilon)\frac{\partial \varepsilon^k}{\partial \varepsilon} \,\mathrm{d}\varepsilon,\end{aligned}$$

using integration by parts. We notice that $P(\varepsilon)$ vanishes for $\varepsilon > \mathrm{Amp}(f)$. We can therefore substitute $\mathrm{Amp}(f)$ for ∞ and get

$$\mathrm{Pers}_k(f,t) \quad \leq \quad t^k P(t) + k \int_{\varepsilon=t}^{\mathrm{Amp}(f)} P(\varepsilon)\varepsilon^{k-1} \,\mathrm{d}\varepsilon.$$

The claimed inequality follows. □

Polynomial growth and bounded total persistence. Assume now that the size of the smallest triangulation *grows polynomially* with one over the mesh. By this we mean there are constants C_0 and M such that $N(r) \leq C_0/r^M$ for every $r > 0$. For example such a behavior holds for any bilipschitz image of an M-dimensional Euclidean simplicial complex, since we can obtain small triangulations as images of subdivisions of the simplicial complex. This general case includes in particular compact finite-dimensional Riemannian manifolds. Let now A and B be the two terms of the upper bound in the Moment Lemma, that is, $\mathrm{Pers}_k(f,t) \leq A + B$. Assuming polynomial growth and setting $k = M + \delta$ for some constant $\delta > 0$ we can find constant upper bounds for both terms,

$$\begin{aligned}A &\leq t^{M+\delta} C_0 \frac{\mathrm{Lip}(f)^M}{t^M} \\ &\leq C_0 \mathrm{Lip}(f)^M \mathrm{Amp}(f)^\delta;\end{aligned}$$

$$\begin{aligned} B &\leq (M+\delta)\int_{\varepsilon=t}^{\mathrm{Amp}(f)} C_0 \frac{\mathrm{Lip}(f)^M}{\varepsilon^M}\varepsilon^{M+\delta-1}\,\mathrm{d}\varepsilon \\ &= C_0\mathrm{Lip}(f)^M(M+\delta)\left[\frac{1}{\delta}\varepsilon^\delta\right]_{\varepsilon=t}^{\mathrm{Amp}(f)} \\ &\leq C_0\mathrm{Lip}(f)^M\mathrm{Amp}(f)^\delta\frac{M+\delta}{\delta}. \end{aligned}$$

Both constant upper bounds depend only on the space $\mathbb{X}$, the Lipschitz constant of the function f, and the chosen constant δ. This result motivates the introduction of the following concept.

DEFINITION. A metric space $\mathbb{X}$ *implies bounded degree-k total persistence* if there is a constant $C_{\mathbb{X}}$ that depends only on $\mathbb{X}$ such that $\mathrm{Pers}_k(f) \leq C_{\mathbb{X}}$ for every tame function $f : \mathbb{X} \to \mathbb{R}$ with Lipschitz constant $\mathrm{Lip}(f) \leq 1$.

For general Lipschitz functions f we get $\mathrm{Pers}_k(f) \leq C$, where $C = C_{\mathbb{X}}\mathrm{Lip}(f)^k$. Consider $\mathbb{X} = \mathbb{S}^n$ as an example and note that $N(r) \leq C_0/r^n$ implies $\mathrm{Pers}_k(f) \leq C$ for every $k = n + \delta$, where $\delta > 0$. This gives $k = 1 + \delta$ for the circle, $\mathbb{S}^1$, but we will see later that in this special case the exponent can be lowered to $k = 1$.

3 Results

The basic intuition for our results is that for a Lipschitz function on a compact space to have a large number of homological critical values their persistence must be small.

Wasserstein distance. The first way of making this intuition precise uses the L_∞-stability of tame functions proved in [3] and proves the L_p-stability of tame Lipschitz functions on compact metric spaces. Let $f, g : \mathbb{X} \to \mathbb{R}$ be two tame functions with persistence diagrams $\mathrm{Dgm}_\ell(f)$ and $\mathrm{Dgm}_\ell(g)$, one for each dimension ℓ. Assuming bounded degree-k total persistence, we prove that the sum of p-th powers of distances between matched points in the two diagrams is stable for every $p > k$. Specifically, we define the *degree-p Wasserstein distance* between the persistence diagrams of f and g,

$$W_p(f,g) \quad = \quad \left[\sum_\ell \inf_{\gamma_\ell}\sum_x \|x-\gamma_\ell(x)\|_\infty^p\right]^{\frac{1}{p}},$$

where the first sum is over all dimensions ℓ, the infimum is over all bijections $\gamma_\ell : \mathrm{Dgm}_\ell(f) \to \mathrm{Dgm}_\ell(g)$, and the second sum is over all points x in $\mathrm{Dgm}_\ell(f)$ [14]. This notion of distance between distributions is popular in computer vision; see also the Monge-Kantorovich transportation problem [10, 11].

WASSERSTEIN STABILITY THEOREM. Let $\mathbb{X}$ be a triangulable, compact metric space that implies bounded degree-k total persistence, for $k \geq 1$, and let $f, g : \mathbb{X} \to \mathbb{R}$ be two tame Lipschitz functions. Then

$$W_p(f,g) \quad \leq \quad C^{\frac{1}{p}} \cdot \|f-g\|_\infty^{1-\frac{k}{p}}$$

for all $p \geq k$, where $C = C_{\mathbb{X}} \max\{\mathrm{Lip}(f)^k, \mathrm{Lip}(g)^k\}$.

PROOF. Let $\gamma_\ell : \mathrm{Dgm}_\ell(f) \to \mathrm{Dgm}_\ell(g)$ be the bijection that realizes the bottleneck distance, that is, $\|x-\gamma_\ell(x)\|_\infty \leq \varepsilon = \|f-g\|_\infty$ for each point x in the first diagram. In addition, we require that $\|x-\gamma_\ell(x)\|_\infty \leq \frac{1}{2}[\mathrm{pers}(x) + \mathrm{pers}(\gamma_\ell(x))]$. Indeed, if this inequality does not hold then $\mathrm{pers}(x) \leq 2\varepsilon$ and $\mathrm{pers}(\gamma_\ell(x)) \leq 2\varepsilon$ and we can change the bijection by matching both points with points on the diagonal within L_∞-distance ε. The p-th power of the degree-p Wasserstein distance therefore satisfies

$$\begin{aligned} W_p(f,g)^p &\leq \sum_{\ell,x}\|x-\gamma_\ell(x)\|_\infty^p \\ &\leq \varepsilon^{p-k}\sum_{\ell,x}\|x-\gamma_\ell(x)\|_\infty^k \\ &\leq \frac{\varepsilon^{p-k}}{2^k}\sum_{\ell,x}[\mathrm{pers}(x)+\mathrm{pers}(\gamma_\ell(x))]^k. \end{aligned}$$

Recall that every convex function f satisfies $f(\frac{u+v}{2}) \leq \frac{1}{2}[f(u)+f(v)]$. Noting that taking the k-th power is convex, we set $u = 2\mathrm{pers}(x)$ and $v = 2\mathrm{pers}(\gamma_\ell(x))$ to further bound $W_p(f,g)^p$ from above by

$$\frac{\varepsilon^{p-k}}{2^k}\sum_{\ell,x}[(2\mathrm{pers}(x))^k + (2\mathrm{pers}(\gamma_\ell(x)))^k].$$

Using the assumption that $\mathbb{X}$ implies bounded degree-k total persistence, we bound this sum by $C\varepsilon^{p-k}$, as required. □

Total persistence. Recall that the degree-k total persistence of a functions $f : \mathbb{X} \to \mathbb{R}$ describes f by a single number. We prove that for Lipschitz functions on triangulable, compact metric spaces this description is stable. Assuming bounded degree-k total persistence, we prove the degree-p total persistence is stable for $p \geq k+1$.

TOTAL PERSISTENCE STABILITY THEOREM. Let $\mathbb{X}$ be a triangulable, compact metric space that implies degree-k total persistence for $k \geq 0$, and let $f, g : \mathbb{X} \to \mathbb{R}$ be two tame Lipschitz functions. Then

$$|\mathrm{Pers}_p(f) - \mathrm{Pers}_p(g)| \quad \leq \quad 4pw^{p-1-k}C \cdot \|f-g\|_\infty$$

for every real $p \geq k+1$, where the constant C is equal to $C_{\mathbb{X}}\max\{\mathrm{Lip}(f)^k, \mathrm{Lip}(g)^k\}$ and w is bounded from above by $\max\{\mathrm{Amp}(f), \mathrm{Amp}(g)\}$.

PROOF. We begin by noting that $y^p - x^p = \int_x^y pz^{p-1}\,\mathrm{d}z \leq p|y-x|\max\{x,y\}^{p-1}$ for all $x, y \geq 0$ and $p \geq 1$. We use the Stability Theorem in [3] to index the persistences of the points in the diagrams of f and g such that

$$\begin{aligned} \mathrm{Pers}_k(f) &= p_1^k + p_2^k + \ldots + p_m^k \leq C; \\ \mathrm{Pers}_k(g) &= q_1^k + q_2^k + \ldots + q_m^k \leq C, \end{aligned}$$

and $|p_i - q_i| \leq 2\varepsilon$ for all i, where $\varepsilon = \|f - g\|_\infty$, possibly after adding zeroes. Now let w be the maximum of the p_i and the q_i. Writing $\Delta = \mathrm{Pers}_p(f) - \mathrm{Pers}_p(g)$ we get

$$\begin{aligned} |\Delta| &\leq \sum_{i=1}^{m} |p_i^p - q_i^p| \\ &\leq \sum_{i=1}^{m} p|p_i - q_i| \max\{p_i, q_i\}^{p-1} \\ &\leq p(2\varepsilon)(2C)w^{p-1-k}, \end{aligned}$$

as claimed. □

4 Applications

As mentioned in Section 1, the two results in Section 3 are motivated by applications to the analysis of gene expression data in development. To change the level of expression needs work, namely the production of RNA to grow and the degradation of RNA to shrink. It is thus natural to assume the functions that expression in time are Lipschitz. In the first application, we use a Lipschitz function on the interval, $f_W : [0, 1] \to \mathbb{R}$, and in the second a Lipschitz function on the circle, $f_T : \mathbb{S}^1 \to \mathbb{R}$. For both we get bounded degree-1 total persistence which is slightly better than what we get from the Moment Lemma. In the case of the interval, the constant upper bound follows from the observation that the sum of persistences is at most the total variation,

$$\mathrm{Pers}_1(f_W) \leq \int_{s=0}^{1} |f'_W(s)| \, \mathrm{d}s.$$

Because f_W is Lipschitz, this integral is at most $\mathrm{Lip}(f_W)$. In the case of the circle, the sum of persistences is half the total variation and we get $\mathrm{Pers}_1(f_T) \leq \pi \mathrm{Lip}(f_T)$.

Similarity in gene expression. We use the result about functions on the unit interval to define a similarity measure for gene expression patterns to be used in clustering. The biological question concerns the expression of genes along the root of the plant arabidopsis [1]. Concrete insights are sought by analyzing the recently obtained microarray data for about 20,000 genes expressed in 14 cell types over 13 stages in the development of the plant arabidopsis [2]. We refer to the stages as longitudinal data since it is taken along the root so that position corresponds to the age of the cell. For each cell type and each gene we have a sequence of 13 measurements which we interpret as samples of a function $f : [0, 1] \to \mathbb{R}$. The first step in discovering gene regulation and cellular communication from this data is to cluster the genes according to their expression patterns. Genes that are expressed in the same longitudinal pattern may be co-regulated or may work together to regulate other genes. Using the clusters we form hypotheses that can then be scrutinized. It is important to keep in mind that the microarray data is sparse and noisy and the longitudinal slices do not accurately represent actual time. We therefore look for qualitative similarity rather than for agreement in detail.

There are many clustering methods that are traditionally applied to expression data. All of these methods require a metric and most are easily adapted to accommodate different choices. Commonly used are the Euclidean metric, the Pearson correlation, and simple boolean metrics. However, the Wasserstein distance between the persistence diagrams provides an alternative that we believe is more appropriate in some situations. Specifically, we suggest using $W_2(f, g)$. By design, it adapts to local variation without requiring that change rigidly occurs at the same stage. Furthermore, this measure is stable. Indeed, the Wasserstein Stability Theorem together with $\mathrm{Pers}_1(f) \leq C$ and $\mathrm{Pers}_1(g) \leq C$ implies $W_2(f, g) \leq 2C^{\frac{1}{2}} \cdot \|f - g\|_\infty^{\frac{1}{2}}$.

Rhythmic gene expression. The background for the second application is the development of somites in vertebrates. Specifically, we consider the mouse embryo in which usually 65 somites are formed in a rhythmic process that takes about two hours per somite. They provide the basic body structure for the organism. The work of Pourquié [13] reduces this rhythm to the periodic expression of genes. In an effort to expand the early results, Dequéant and Pourquié used a microarray experiment to collect data on the expression of about 7,500 genes at 17 time-points within a single period. After finding twenty-seven genes involved in the rhythmic process [5], the task is now to complete the picture by identifying additional players needed to uncover the biological clocks that drive the somite development. The task is made difficult by the sparsity of the data and the noise.

To be specific, we think of the 17 measurements as ordered samples of a function $f : \mathbb{S}^1 \to \mathbb{R}$ and base our assessment of periodicity on counting critical points. More precisely, we integrate the number of critical points over the ε-*simplifications* $f_\varepsilon : \mathbb{S}^1 \to \mathbb{R}$ defined such that $\|f - f_\varepsilon\|_\infty \leq \varepsilon$ and the diagrams of f_ε agree with those of f except that they contain no points of persistence ε or less; see [8]. For functions on the circle it is easy to see that f_ε exists for every value of ε. To simplify notation, let us assume that the amplitude of f is one. Writing $c_0(f)$ for the number of minima and $c_1(f)$ for the number of maxima, the sequence of measures is defined inductively as

$$\begin{aligned} M_0(f) &= \frac{c_0(f) + c_1(f)}{2}; \\ M_i(f) &= \int_{\varepsilon=0}^{1} M_{i-1}(f_\varepsilon) \, \mathrm{d}\varepsilon, \end{aligned}$$

for $i \geq 1$. For example, M_1 integrates half the number of critical points over all ε-simplifications. Two critical points paired by persistence contribute to the integral for ε from zero to their persistence, which implies that $M_1(f)$ is equal to the sum of persistences. As noted before, this is half the

total variation of f. More generally, $M_i(f) = \sum_x \mathrm{pers}(x)^i$, where the sum is over all points x in $\mathrm{Dgm}_0(f)$. ¿From our analysis we know that M_i is stable for $i \geq 2$ and it is not difficult to prove that it is not stable for $i \leq 1$. In the ranking of functions we indeed see a marked change in the distribution of the twenty-seven validated genes when we go from M_1 to M_2. For $i = 0, 1$ there are relatively few near the top, for $i = 2$ this number increases dramatically, and for $i > 2$ that number stays about the same. We refer to [6] for details of the experiments and the comparison of M_2 with other assessments of periodicity.

5 Discussion

Recall that the size of the smallest triangulation of $\mathbb{S}^n$ grows linearly with the n-th power of one over the mesh, $N(r) \leq C_0/r^n$. It follows from the Moment Lemma that $\mathbb{S}^n$ implies bounded degree-k total persistence for $k = n + \delta$, for every constant $\delta > 0$. It is also easy to see that $k \geq n$ is necessary for we can populate $\mathbb{S}^n$ with $1/r^n$ cones each supported by a spherical cap of radius r and contributing a point of persistence r. It would be interesting to know whether $\mathbb{S}^n$ implies bounded degree-k total persistence for $k = n$. In Section 4, we proved that this is the case for $n = 1$ using an argument that circumvents the use of a triangulation. The authors of this paper generalized this proof to $n = 2$ but the question is still open for $n \geq 3$.

Acknowledgments

The authors thank Olivier Pourquié and Philip Benfey for motivating the work reported in this paper by their gene expression experiments and Dmitriy Morozov for insightful technical discussions. They also thank two anonymous referees for valuable comments that helped improve the presentation but also the results in this paper.

References

[1] P. N. BENFEY AND B. SCHERES. Root development. *J. Current Biology* **10** (2000), R813–815.

[2] S. M. BRADY, D. A. ORLANDO, J.-Y. LEE, J. Y. WANG, J. KOCH, J. R. DINNENY, D. MACE, U. OHLER AND P. N. BENFEY. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318** (2007), 801–806.

[3] D. COHEN-STEINER, H. EDELSBRUNNER AND J. HARER. Stability of persistence diagrams. *Discrete Comput. Geom.* **37** (2007), 103–120.

[4] D. COHEN-STEINER, H. EDELSBRUNNER AND J. HARER. Extending persistence using Poincaré and Lefschetz duality. *Found. Comput. Math.*, to appear.

[5] M.-L. DEQUÉANT, E. GLYNN, K. GAUDENZ, M. WAHL, J. CHEN, A. MUSHEGIAN AND O. POURQUIÉ. A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science* **314** (2006), 1595–1598.

[6] M.-L. DEQUÉANT, S. AHNERT, H. EDELSBRUNNER, T. M. A. FINK, E. F. GLYNN, G. HATTEM, A. KUDLICKI, Y. MILEYKO, J. MORTON, A. R. MUSHEGIAN, L. PACHTER, M. ROWICKA, A. SHIU, B. STURMFELS AND O. POURQUIÉ. Comparison of pattern detection methods in microarray time series of the segmentation clock. *PLoS ONE* **3** (2008), e2856, doi:10.1371/journal.pone.-0002856.

[7] H. EDELSBRUNNER, D. LETSCHER AND A. ZOMORODIAN. Topological persistence and simplification. *Discrete Comput. Geom.* **28** (2002), 511–533.

[8] H. EDELSBRUNNER, D. MOROZOV AND V. PASCUCCI. Persistence-sensitive simplification of functions on 2-manifolds. *In* "Proc. 22th Ann. Sympos. Comput. Geom., 2006", 127–134.

[9] P. FROSINI AND C. LANDI. Size theory as a topological tool for computer vision. *Pattern Recognition and Image Analysis* **9** (1999), 596–603.

[10] L. V. KANTOROVICH. On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS* **37** (1942), 199–226.

[11] G. MONGE. Mémoire sur la théorie des déblais et des remblais. In *Histoire de l'Académie Royale des Sciences de Paris* (1781), 666–704.

[12] J. R. MUNKRES. *Elements of Algebraic Topology.* Addison-Wesley, Redwood City, California, 1984.

[13] O. POURQUIÉ. The segmentation clock: converting embryonic time into spatial pattern. *Science* **301** (2003), 328–330.

[14] L. N. WASSERSTEIN. Markov processes over denumerable products of spaces describing large systems of automata. *Problems of Information Transmission* **5** (1969), 47–52.

[15] A. ZOMORODIAN AND G. CARLSSON. Computing persistent homology. *Discrete Comput. Geom.* **33** (2005), 249–274.