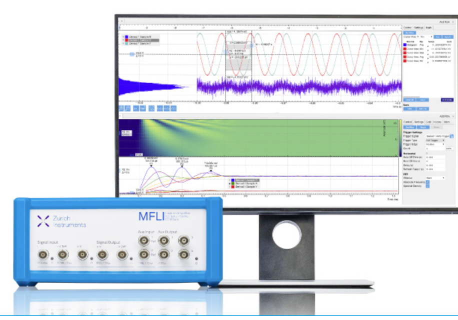
Strong topology on the set of persistence diagrams

Cite as: AIP Conference Proceedings **2164**, 040006 (2019); https://doi.org/10.1063/1.5130798
Published Online: 24 October 2019

M. Zarichnyi, A. Savchenko, and V. Kiosak



Strong Topology on the Set of Persistence Diagrams

M. Zarichnyi[1,a)], A. Savchenko[2,b)] and V. Kiosak[3,c)]

[1]*Faculty of Mathematics and Natural Sciences, University of Rzeszów, 1 Prof. St. Pigoń str. 35-310 Rzeszów, Poland*
[2]*Kherson State Agrarian University, 23 Stretenska str., 73006 Kherson, Ukraine*
[3]*Institute of Engineering Odessa State Academy of Civil Engineering and Architecture 4 Didrihson str., 65029 Odessa, Ukraine*

[a)]zarichnyi@yahoo.com
[b)]savchenko.o.g@ukr.net
[c)]kiosakv@ukr.net

Abstract. We endow the set of persistence diagrams with the strong topology (the topology of countable direct limit of increasing sequence of bounded subsets considered in the bottleneck distance). The topology of the obtained space is described. Also, we prove that the space of persistence diagrams with the bottleneck metric has infinite asymptotic dimension in the sense of Gromov.

INTRODUCTION

Topological Data Analysis (TDA) is a field in applied mathematics concentrated around investigation of big data by topological methods. Imposing metric structures in the data set allows for applying techniques from algebraic topology. In this way, the notion of persistent homology was introduced [4]. The persistent homology plays an important role in TDA. The persistence diagrams are used to characterize persistent homology and thus to describe geometric properties of data. The set of all persistence diagrams can be endowed with different metrics. The most known are the Wasserstein metric and bottleneck metric. The spaces of persistence diagrams are object of considerations in numerous publications (see, *e.g.*, [5, 9, 13, 14, 17]). In particular, in [14] a characterization theorem for compact subsets in the space of persistence diagrams is proved. It is proved in [2] that the space of persistent diagrams is of infinite asymptotic dimension in the sense of Gromov. This concerns the Wasserstein metric on the set of persistence diagrams. Answering a question from [2] we prove an analogous result for the bottleneck metric on this set. As we remark below, the set of all persistence diagrams is nothing but the infinite symmetric power of the upper (positive) half-plane. In this note we consider the strong (direct limit) topology on this set. One of our results is that the space of the persistence diagrams with this topology is homeomorphic to the countable direct limit of the euclidean spaces.

PRELIMINARIES

Persistence diagrams

Let $\Delta = \{(x, y) \in \mathbb{R}^2_+ \mid x = y\}$, $\hat{X} = \{(x, y) \in \mathbb{R}^2_+ \mid x \le y\}$, and $X = \hat{X} \setminus \Delta$. For any $n \in \mathbb{N}$, let $\hat{X}_n = \{(x, y) \in \hat{X} \mid y \le n\}$, $X_n = \hat{X}_n \setminus \Delta$. A *persistence diagram* is a function $\mu\colon X \to \mathbb{Z}_+$ such that $\mu(a) = 0$ for all but finitely many $a \in \hat{X}$ and $\mu(a) = 0$ for all $a \in \Delta$. The *support* of μ is the set $\operatorname{supp}(\mu) = \{a \in \hat{X} \mid \mu(a) > 0\}$. By $\mathcal{D}$ we denote the set of all persistence diagrams. Given $n \in \mathbb{N}$, we denote by $\mathcal{D}_n$ the set of all $\mu \in \mathcal{D}$ such that $|\operatorname{supp}(\mu)| \le n$.

Bottleneck distance

Let $\mu \in \mathcal{D}$. A *sequential representation* of μ is a finite sequence $(a_1, \dots, a_k)$ such that the following are satisfied:

1. for every $a \in \operatorname{supp}(\mu)$, $|\{i \le m \mid a = a_i\}| = \mu(a)$;

Application of Mathematics in Technical and Natural Sciences
AIP Conf. Proc. 2164, 040006-1–040006-4; https://doi.org/10.1063/1.5130798
Published by AIP Publishing. 978-0-7354-1909-4/$30.00

2. if $a_i \notin \mathrm{supp}(\mu)$, then $a_i \in \Delta$.

The number k is said to be the *length* of the representation $(a_1, \dots, a_k)$. By S_k, the group of permutations of the set $\{1, \dots, k\}$ is denoted. Let $\mu, \nu \in \mathcal{D}$. We define

$$d(\mu, \nu) = \inf\{\min\{\max\{\rho(a_i, b_{\sigma(i)}) \mid 1 \le i \le k\} \mid \sigma \in S_k\} \mid (a_1, \dots, a_k),\ (b_1, \dots, b_k) \text{ are sequential representations of } \mu \text{ and } \nu \text{ respectively},\ k \in \mathbb{N}\}.$$

(The assignment $a_i \mapsto b_{\sigma(i)}$, $i = 1, \dots, k$, is said to be a *matching* (see, *e.g.*, [6] for details). The function d is known to be a metric on $\mathcal{D}$ (the bottleneck metric; see, *e.g.*, [6]).

Space $\mathbb{R}^\infty$

Recall that the direct limit of the increasing sequence of topological spaces $X_1 \subset X_2 \subset \dots$ (here X_n is a subspace of X_{n+1}, for each n) is the set $X = \cup_{n=1}^{\infty} X_n$ endowed with the strongest topology inducing the original topology on each X_n. The obtained topological space is denoted by $\varinjlim X_n$. We identify every $(x_1, \dots, x_n) \in \mathbb{R}^n$ with $(x_1, \dots, x_n, 0) \in \mathbb{R}^{n+1}$. Thus, $\mathbb{R}^n$ is regarded as a subspace in $\mathbb{R}^{n+1}$. We denote by $\mathbb{R}^\infty$ the direct limit of the sequence $\mathbb{R} \subset \mathbb{R}^2 \subset \mathbb{R}^3 \subset \dots$. A characterization theorem for the space $\mathbb{R}^\infty$ is proved by K. Sakai [15].

Theorem 1. (Characterization Theorem for $\mathbb{R}^\infty$) *Let X be a countable direct limit of finite-dimensional compact metrizable spaces. The following are equivalent.*

1. *X is homeomorphic to $\mathbb{R}^\infty$;*
2. *for every finite-dimensional compact metrizable pair (A, B) and every embedding $f\colon B \to X$ there exists an embedding $\bar{f}\colon A \to X$ that extends f.*

Asymptotic dimension

Let Y be a metric space. A family $\mathcal{A}$ of subsets of X is said to be *uniformly bounded* if $\sup\{\mathrm{diam}(A) \mid A \in \mathcal{A}\} < \infty$. Given $D > 0$, we say that a family $\mathcal{A}$ is *D-disjoint* if, for every distinct $A, B \in \mathcal{A}$, $\mathrm{dist}(A, B) \ge D$ We say that the asymptotic dimension of Y does not exceed n, if, for any $D > 0$, there exists a uniformly bounded cover $\mathcal{U}$ of Y such that $\mathcal{U} = \cup_{i=0}^{n} \mathcal{U}_i$, where every family $\mathcal{U}_i$ is D-disjoint, $i = 0, \dots, n$. The notion of asymptotic dimension is defined by M. Gromov [8]. See, *e.g.*, [1] for properties of the asymptotic dimension.

MAIN RESULTS

Strong topology on the space of persistence diagrams

For every $n \in \mathbb{N}$, let

$$\mathcal{D}_n = \{\mu \in \mathcal{D} \mid |\mathrm{supp}(\mu)| \le n \text{ and } \mathrm{supp}(\mu) \subset X_n\}$$

and $\mathcal{D}_\infty = \varinjlim \mathcal{D}_n$.

Theorem 2. *The space $\mathcal{D}_\infty$ is homeomorphic to $\mathbb{R}^\infty$.*

Proof. For any $n \in \mathbb{N}$, define a map $\xi_n\colon \hat{X}_n^n \to \mathcal{D}_n$ as follows: $\xi_n(a_1, \dots, a_n)(a) = |\{i \mid a = a_i\}|$, for all $a \in X_n$. Note that this map is clearly continuous and it admits a factorization $\xi_n = \xi'_n \xi''_n$, where $\xi''\colon \hat{X}_n^n \to (\hat{X}_n/(\hat{X}_n \cap \Delta)_n)^n$ is the factorization map, where $*_n$ stands for $\hat{X}_n \cap \Delta$ (actually, $\xi''_n = q^n$, where $q\colon \hat{X}^n \to \hat{X}_n/(\hat{X}_n \cap \Delta)_n$ is the factorization map). Therefore, $\mathcal{D}_n$ is the orbit space of the action of the group S_n on the space $(\hat{X}_n/(\hat{X}_n \cap \Delta))^n$ by permutation of coordinates. In other words, $\mathcal{D}_n$ is homeomorphic to the nth symmetric power $SP^n(\hat{X}_n/(\hat{X}_n \cap \Delta))$. The orbit containing $(x_1, \dots, x_n)$ will be denoted by $[x_1, \dots, x_n]$. We denote by $*_n$ the point $q(\hat{X}_n \cap \Delta) \in \hat{X}_n/(\hat{X}_n \cap \Delta)$. Identifying $*_n$ with $*_{n+1}$, one can consider $\hat{X}_n/(\hat{X}_n \cap \Delta)$ as a subset of $\hat{X}_{n+1}/(\hat{X}_{n+1} \cap \Delta)$. Then identifying $[x_1, \dots, x_n] \in SP^n(\hat{X}_n/(\hat{X}_n \cap \Delta))$ with $[x_1, \dots, x_n, *_{n+1}] \in SP^{n+1}(\hat{X}_{n+1}/(\hat{X}_{n+1} \cap \Delta))$ we finally obtain that $\mathcal{D}_\infty$ is homeomorphic to the space

$$\varinjlim SP^n(\hat{X}_n/(\hat{X}_n \cap \Delta)) = SP^\infty(\varinjlim \hat{X}_n/(\hat{X}_n \cap \Delta)).$$

The latter space is known as the infinite symmetric power construction [7]. For every $n \in \mathbb{N}$, the space $\hat{X}_n/(\hat{X}_n \cap \Delta)$ is homeomorphic to the 2-dimensional disc. Therefore, the space $SP^n(\hat{X}_n/(\hat{X}_n \cap \Delta))$ is a finite dimensional absolute retract (AR), see [18]. Similarly as in [19] one can prove that the space $SP^\infty(\varinjlim \hat{X}_n/(\hat{X}_n \cap \Delta)$ (and also the space $\mathcal{D}_\infty$) is homeomorphic to $\mathbb{R}^\infty$. For the sake of completeness, we provide here a proof based on Sakai's Characterization Theorem 1. Let (A, B) be a finite-dimensional compact metrizable pair an let $f: B \to \mathcal{D}_\infty$ be an embedding. Then there exists $n \in \mathbb{N}$ such that $f(B) \subset \mathcal{D}_n$. As we already remarked, $\mathcal{D}_n$ is an absolute retract. Thus, there exists an extension $g: A \to \mathcal{D}_n$ of f. We denote by $\alpha: A \to A/B$ the quotient map. Since the quotient space A/B is finite-dimensional, there exists an embedding $i: A/B \to [0,1]^m$, for some m. Without loss of generality we assume that $\alpha(A/B) = 0$, $\alpha(A \setminus B) \subset (0,1)^m$, and $m > n$. Write $i(a) = (i_1(a), \dots, i_m(a))$. Now, given $x \in A$, write $g(x)$ as $[g_1(x), \dots, g_n(x)]$. Then define

$$\begin{aligned}\bar{f}(x) =&[g_1(x), \dots, g_n(x),\\ &(1, 1 + i_1(\alpha(x))), \dots, (2n+1, 2n+1+i_1(\alpha(x))),\\ &(2n+2, 2n+2+i_2(\alpha(x)), \dots, (4n+3, 4n+3+i_2(\alpha(x))),\\ &\dots\\ &((m-1)(2n+1)+1, (m-1)(2n+1)+1+i_m(\alpha(x))), \dots,\\ &(m(2n+1)+1, m(2n+1)+1+i_m(\alpha(x)))].\end{aligned}$$

First, note that $\bar{f}$ is well defined and continuous. If $x \in B$, then $i_k(\alpha(x)) = 0$, $k = 1, \dots, n(2n+1)$, and therefore

$$\bar{f}(x) = [g_1(x), \dots, g_n(x), (1,1), \dots, (n(2n+1), n(2n+1)))] = f(x),$$

because of our identifications. Clearly, $\bar{f}$ is well-defined and continuous. Since A is compact, in order to prove that $\bar{f}$ is an embedding it is sufficient to prove that $\bar{f}$ is injective. Let $x, y \in A$, $x \neq y$. If $x \in B$ and $y \in A \setminus B$, then $|\mathrm{supp}(\bar{f}(x))| \leq n$ and, since $i_k(\alpha(y)) \neq 0$, for some $k \leq m$, we conclude that $|\mathrm{supp}(\bar{f}(y))| \geq n+1$. If $x, y \in A \setminus B$, then there exists $k \leq m$, such that $i_k(\alpha(x)) \neq i_k(\alpha(y))$. Since $|\mathrm{supp}(g(x))| \leq n$, we see that there exists $j \leq n+1$ such that

$$(j(2n+1)+1, j(2n+1)+1+i_j(\alpha(x)) \in \mathrm{supp}(\bar{f}(x)) \setminus \mathrm{supp}(\bar{f}(y))$$

and therefore $\bar{f}(x) \neq \bar{f}(y)$. The other cases being treated similarly, this proves injectivity of $\bar{f}$. By Theorem 1, $\mathcal{D}_\infty$ is homeomorphic to $\mathbb{R}^\infty$.

Asymptotic dimension of the space of persistence diagrams

We consider the set $\mathcal{D}$ of all persistence diagrams with the bottleneck metric. Given $n \in \mathbb{N}$, consider the set

$$\begin{aligned}K_n =&\{[(n, n+(n+1)+t_1), (2n+1, 2n+1+(n+1)+t_2), \dots,\\ &(n-1)n+(n-2), (n-1)n+(n-2)+(n+1)+t_{n-1})],\\ &(n^2+(n-1), n^2+(n-1)+(n+1)+t_n)] \mid (t_1, \dots, t_n) \in [0,n]^n\}.\end{aligned}$$

Clearly, K_n is isometric to the cube $[0,n]^n$ endowed with the l_∞-metric. This easily follows from the observation that every matching between two points from K_n realizing the bottleneck distance consists of pairs of points from X so that each pair lies on a vertical line. Since n can be chosen arbitrarily large, the known properties of the asymptotic dimension (see, *e.g.*, [1]) imply the following.

Theorem 3. $\quad \mathrm{asdim}\, \mathcal{D} = \infty$.

REMARKS

We conjecture that an analog of Theorem 2 can be proved for the space $\mathcal{D}_\infty = \varinjlim \mathcal{D}_n$ in the case when every $\mathcal{D}_n$ is endowed with the Wasserstein metric (also Wasserstein p, q metric considered in [2]). Since many spaces of persistence diagrams are infinitely-dimensional, one can expect that the methods of infinite-dimensional topology, in particular, the theory of infinite-dimensional manifolds, will be useful in their investigations. In [5], the space

$\mathcal{D}^b_N$ of bounded persistent diagrams with less than N points is mentioned. We consider the space $\tilde{\mathcal{D}}_N$ of exactly N points (taking into account the multiplicities), $N \in \mathbb{N}$. Having in mind the mentioned identification of persistence diagrams and symmetric powers one can derive from [18, Theorem 4.5] that the space $\tilde{\mathcal{D}}_N$ is homeomorphic to the euclidean space $\mathbb{R}^{2N}$. This leads to the question of description of topology of the subspace $\tilde{\mathcal{D}}_{\leq N} = \cup_{i\leq N}\tilde{\mathcal{D}}_i$ of $\mathcal{D}$. In this note we restricted ourselves with persistence diagrams of finite support. In some publications, persistence diagrams with countably many points are also considered. In particular, it is known that the latter spaces are complete in the Wasserstein metric. The possible applications could be sought in the [10, 11, 12, 16]. In the subsequent publications we are going to consider the geometry of the complete spaces of persistence diagrams and some of their subspaces.

REFERENCES

[1] G.Bell and A.Dranishnikov (2008) Asymptotic dimension, *Topol. Appl.* **155**, 1265–1296.

[2] G. Bell, A. Lawson, C. Neil Pritchard, and D. Yasaki, The space of persistence diagrams has infinite asymptotic dimension, arXiv:1902.02288.

[3] D. Burago, Yu. Burago, and S. Ivanov, *A Course in Metric Geometry*, Vol. 33 (AMS GSM, 2001).

[4] G. Carlsson (2009) Topology and data, *Bulletin of the AMS* **46**, 255–308.

[5] M. Carrière, M. Cuturi, and S. Oudot, Sliced Wasserstein kernel for persistence diagrams, arXiv:1706.03358.

[6] F. Chazal, V. de Silva, M. Glisse, and S. Oudot, *The Structure and Stability of Persistence Modules*, Springer Briefs in Mathematics (Springer, Cham, 2016).

[7] A. Dold and R. Thom (1958) Quasifaserungen und unendliche symmetrische Produkte, *Annals of Mathematics, Second Series* **67**, 239–281.

[8] M. Gromov, "Asymptotic invariants of infinite groups," in *Geometric Group Theory* **2**, Sussex, (1991), *London Math. Soc. Lecture Note Ser.* (Cambridge Univ. Press, Cambridge, 1993), **182**, 1–295.

[9] C. Li, M. Ovsjanikov, and F. Chazal, "Persistence-based structural recognition," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 1995–2002.

[10] V. Kiosak (2012) On the conformal mappings of quasi-Einstein spaces, *Journal of Mathematical Sciences* **184**(1), 12–18.

[11] V. Kiosak, O. Lesechko, and O. Savchenko, "Mappings of spaces with affine connection," in *17th Conference on Applied Mathematics APLIMAT 2018* - Procedings (Bratislava, 2018), pp. 563–569.

[12] V. Kiosak, O. Savchenko, and T. Shevchenko, "Holomorphically Projective Mappings of Special Kahler Manifolds," in *AMiTaNS'18* AIP CP**2025** edited by M.D. Todorov (American Institute of Physics, Melville, NY, 2018), paper 08004.

[13] Y. Mileyko, S. Mukherjee, and J. Harer (2011) Probability measures on the space of persistence diagrams, *Inverse Problems* **27**(12), 124007 22p.

[14] J. A. Perea, E. Munch, and F. A. Khasawneh, Approximating continuous functions on persistence diagrams using template functions, arXiv:1902.07190.

[15] K. Sakai (1984) On $\mathbf{R}^\infty$ and Q^∞-manifolds", *Topol. Appl.* **8**(1), 69–79.

[16] A. Savchenko and M. Zarichnyi (2010) Metrization of free groups on ultrametrics spaces, *Topology and its Applications* **157**(4), 724–729.

[17] K. Turner and G. Spreemann, Same but different: distance correlations between topological summaries, arXiv:1903.01051.

[18] C. H. Wagner, *Symmetric, Cyclic, and Permutation Products of Manifolds* (Instytut Matematyczny Polskiej Akademi Nauk, Warszawa, 1980).

[19] M. M. Zarichnyi (1982) Free topological groups of absolute neighborhood retracts and infinite-dimensional manifolds, *Dokl. Akad. Nauk SSSR* **266**(3), 541–544. [in Russian]