



HAL
open science

On the choice of weight functions for linear representations of persistence diagrams

Vincent Divol, Wolfgang Polonik

► **To cite this version:**

Vincent Divol, Wolfgang Polonik. On the choice of weight functions for linear representations of persistence diagrams. *Journal of Applied and Computational Topology*, 2019, 10.1007/s41468-019-00032-z . hal-01833660v2

HAL Id: hal-01833660

<https://inria.hal.science/hal-01833660v2>

Submitted on 25 Apr 2019 (v2), last revised 11 Nov 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON THE CHOICE OF WEIGHT FUNCTIONS FOR LINEAR REPRESENTATIONS OF PERSISTENCE DIAGRAMS

VINCENT DIVOL AND WOLFGANG POLONIK

ABSTRACT. Persistence diagrams are efficient descriptors of the topology of a point cloud. As they do not naturally belong to a Hilbert space, standard statistical methods cannot be directly applied to them. Instead, feature maps (or representations) are commonly used for the analysis. A large class of feature maps, which we call *linear*, depends on some weight functions, the choice of which is a critical issue. An important criterion to choose a weight function is to ensure stability of the feature maps with respect to Wasserstein distances on diagrams. We improve known results on the stability of such maps, and extend it to general weight functions. We also address the choice of the weight function by considering an asymptotic setting; assume that \mathbb{X}_n is an i.i.d. sample from a density on $[0, 1]^d$. For the Čech and Rips filtrations, we characterize the weight functions for which the corresponding feature maps converge as n approaches infinity, and by doing so, we prove laws of large numbers for the total persistences of such diagrams. Those two approaches (stability and convergence) lead to the same simple heuristic for tuning weight functions: if the data lies near a d -dimensional manifold, then a sensible choice of weight function is the persistence to the power α with $\alpha \geq d$.

1. INTRODUCTION

Topological data analysis, or TDA (see [13] for a survey) is a recent field at the intersection of computational geometry, statistics and probability theory that has been successfully applied to various scientific areas, including biology [37], chemistry [28], material science [26] or the study of time series [32]. It consists of an array of techniques aimed at understanding the topology of a d -dimensional manifold based on an approximating point cloud \mathbb{X} . For instance, clustering can be seen as the estimation of the connected components of a given manifold. Persistence diagrams are one of the tools used most often in TDA. They are efficient descriptors of the topology of a point cloud, consisting in a multiset D of points in $\mathbb{R}_{>}^2 := \{r = (r_1, r_2) \in \mathbb{R}^2, r_1 < r_2\}$ (see Section 2 for a more precise definition). The space \mathcal{D} of persistence diagrams is not naturally endowed with a Hilbert or Banach space structure, making statistical inference rather awkward. A common scheme to overcome this issue is to use a *representation* or *feature map* $\Phi : \mathcal{D} \rightarrow \mathcal{B}$, where \mathcal{B} is some Banach space: classical machine learning

INRIA PARIS-SACLAY AND UNIVERSITÉ PARIS-SUD

UNIVERSITY OF CALIFORNIA, DAVIS

E-mail addresses: `vincent.divol@inria.fr`, `wpolonik@ucdavis.edu`.

This work was partially supported by the Advanced Grant of the European Research Council GUDHI (Geometric Understanding in Higher Dimensions).

techniques are then applied to $\Phi(D)$ instead of D , where it is assumed that an entire set (or sample) of persistence diagrams is observed. A natural way to create such feature maps is to consider a function $\phi : \mathbb{R}_{>}^2 \rightarrow \mathcal{B}$ and to define

$$(1.1) \quad \Phi(D) := \sum_{r \in D} \phi(r).$$

A multiset can equivalently be seen as a measure. Therefore we let D also denote the measure $\sum_{r \in D} \delta_r$ with δ_r denoting Dirac measure in r . With this notation, $\Phi(D)$ is equal to $D(\phi)$, the integration of ϕ against the measure D . Representations as in (1.1) are called *linear* as they define linear maps from the space of finite signed measures to the Banach space \mathcal{B} . In the following, a representation will always be considered linear. Many linear representations exist in the literature, including persistence surfaces and its variants [1, 14, 25, 31], persistence silhouettes [12] or accumulated persistence function [3]. Notable non-linear representations include persistence landscapes [7], and sliced Wasserstein kernels [8].

In machine learning, a possible way to circumvent the so-called "curse of dimensionality" is to assume that the data lies near some low-dimensional manifold M . Under this assumption, the persistence diagram of the data set (built with the Čech filtration, for instance) is made of two different types of points: points D_{true} far away from the diagonal, which estimate the diagram of the manifold M , and points D_{noise} close to the diagonal, which are generally considered to be "topological noise" (see Figure 1). This interpretation is a consequence of the stability theorem for persistence diagrams; see [15]. If the relevant information lies in the structure of the manifold, then the topological noise indeed represents true noise, and representations of the form $D(\phi)$ are bound to fail if $D_{\text{noise}}(\phi)$ is dominating $D_{\text{true}}(\phi)$. A way to avoid such behaviour is to weigh the points in diagrams by means of a weighting function $w : \mathbb{R}_{>}^2 \rightarrow \mathbb{R}$. If w is chosen properly, i.e. small enough when close to the diagonal, then one can hope that $D_{\text{true}}(w\phi)$ can be separated from $D_{\text{noise}}(w\phi)$. The weight functions w are typically chosen as functions of the persistence $\text{pers}(r) := r_2 - r_1$, a choice which will be made here also. Of course, it is not clear what "small enough" really means, and there are several ways to address the issue.

A first natural answer is to look at the problem from a stability point of view. Indeed, as data are intrinsically noisy, a statistical method has to be stable with respect to some metric in order to be meaningful. Standard metrics on the space of diagrams \mathcal{D} are Wasserstein distances W_p , which under mild assumptions (see [16]) are known to be stable with respect to the data on which diagrams are built. The task therefore becomes to find representations $D(w\phi)$ that are continuous with respect to some Wasserstein distance. Recent work in [25] shows that when sampling from a d -dimensional manifold, a weight function of the form $w(r) = \arctan(A \cdot \text{pers}(r)^\alpha)$ with $\alpha > d + 1$ ensures that a certain class of representations are Lipschitz. Our first contribution is to show that, for a general class of weight functions, a choice of $\alpha > d$ is enough to make *all* linear representations continuous (even Hölderian of exponent $\alpha - d$).

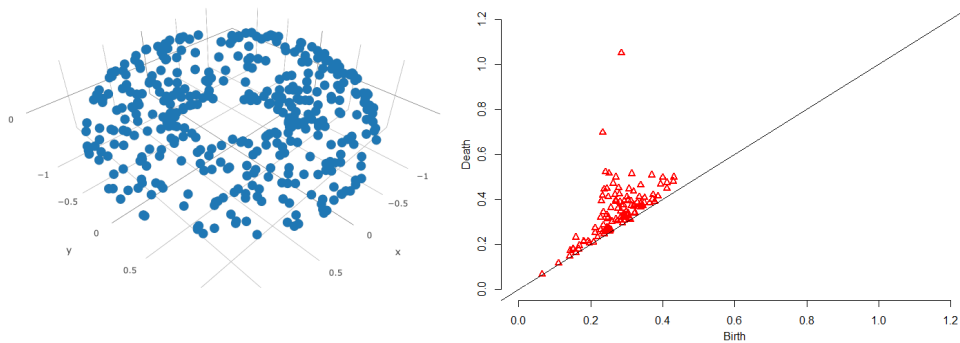


FIGURE 1. The persistence diagram for homology of degree 1 of the Rips filtration of 2000 i.i.d. points uniformly sampled on a torus. There are two distinct points in the diagram, corresponding to the two equivalence classes of one-dimensional holes of the torus.

Our second (and main) contribution is to evaluate closeness to the diagonal from an asymptotic point of view. Assume that a diagram D_n is built on a data set of size n . For which weight functions is $D_{n,\text{noise}}(w\phi)$ none-divergent? Of course, for this question to make sense, a model for the data set has to be specified. A simple model is given by a Poisson (or binomial) process \mathbb{X}_n of intensity n in a cube of dimension d . We denote the corresponding diagrams built on a filtration \mathcal{K} with respect to q -dimensional homology by $\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]$, with \mathcal{K} either the Rips or Čech filtration. A precise definition is given below in Section 2. In this setting, there are no "true" topological features (other than the trivial topological feature of $[0, 1]^d$ being connected), and thus the diagram based on the sampled data is uniquely made of topological noise. A first promising result is the vague convergence of the measure $\mu_q^n := n^{-1} \text{dgm}_q[\mathcal{K}(n^{1/d}\mathbb{X}_n)]$, which was recently proven in [22] for homogeneous Poisson processes in the cube and in [21] for binomial processes on manifolds. However, vague convergence is not enough for our purpose, as neither ϕ nor w have good reasons to have compact support. Our main result, Theorem 4.4 extends result of [21], for processes on the cube, to a stronger convergence, allowing test functions to have both non-compact support (but to converge to 0 near the diagonal) and to have polynomial growth. As a corollary of this general result, the convergence of the α -th total persistence, which plays an important role in TDA, is shown. The α -th total persistence is defined as $\text{Pers}_\alpha(D) := D(\text{pers}^\alpha) = \sum_{r \in D} \text{pers}(r)^\alpha$.

Theorem 1.1. *Let $\alpha > 0$ and let κ be a density on $[0, 1]^d$ such that $0 < \inf \kappa \leq \sup \kappa < \infty$. Let \mathbb{X}_n be either a binomial process with parameters n and κ or a Poisson process of intensity $n\kappa$ in the cube $[0, 1]^d$. Define $\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]$ to be the persistence diagram of \mathbb{X}_n for q -dimensional homology, built with either the Rips or the Čech filtration. Then, with probability one, as $n \rightarrow \infty$*

$$(1.2) \quad n^{\frac{\alpha}{d}-1} \text{Pers}_\alpha(\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]) \rightarrow \mu_q^\kappa(\text{pers}^\alpha) < \infty$$

for some non-degenerate Radon measure μ_q^k on $\mathbb{R}_{>}^2$.

If $D_n := \text{dgm}_q[\mathcal{K}(\mathbb{X}'_n)]$ is built on a point cloud \mathbb{X}'_n of size n on a d -dimensional manifold, one can expect $D_{n,\text{noise}}$ to behave in a similar fashion to that of $\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]$ for \mathbb{X}_n a n -sample on a d -dimensional cube (a manifold looking locally like a cube). Therefore, for $\alpha > 0$, the quantity $D_{n,\text{noise}}(\text{pers}^\alpha)$ should be close to $\text{Pers}_\alpha(\text{dgm}[\mathcal{K}(\mathbb{X}_n)])$, and it can be expected to converge to 0 if and only if the weight function pers^α is such that $\alpha \geq d$. The same heuristic is found through both the approaches (stability and convergence): *a weight function of the form pers^α with $\alpha \geq d$ is sensible if the data lies near a d -dimensional object.*

Further properties of the process $(\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)])_n$ are also shown, namely non-asymptotic rates of decays for the number of points in said diagrams, and the absolute continuity of the marginals of μ_q^k with respect to the Lebesgue measure on \mathbb{R} .

1.1. Related work. Techniques used to derive the large sample results indicated above are closely related to the field of geometric probability, which is the study of geometric quantities arising naturally from point processes in \mathbb{R}^d . A classical result in this field, see [34], proves the convergence of the total length of the minimum spanning tree built on n i.i.d. points in the cube. This pioneering work can be seen as a 0-dimensional special case of our general results about persistence diagrams built for homology of dimension q . This type of result has been extended to a large class of functionals in the works of J. E. Yukich and M. Penrose (see for instance [27, 30, 39] and [29] or [40] for monographs on the subject).

The study of higher dimensional properties of such processes is much more recent. Known results include convergence of Betti numbers for various models and under various asymptotics (see [5, 23, 24, 38]). The paper [4] finds bounds on the persistence of cycles in random complexes, and [22] proves limit theorems for persistence diagrams built on homogeneous point processes. The latter is extended to non-homogeneous processes in [35], and to processes on manifolds in [21]. Note that our results constitute a natural extension of [35]. In [33], higher dimensional analogs of minimum spanning trees, called minimal spanning acycles, were introduced. Minimal spanning acycles exhibits strong links with persistence diagrams and our main theorem can be seen as a convergence result for weighted minimal spanning acycle on geometric random complexes. [33] also proves the convergence of the total 1-persistence for Linial-Meshulam random complexes, which are models of random simplicial complexes of a combinatorial nature rather than a geometric nature.

1.2. Notations.

$\ \cdot\ $	Euclidean distance on \mathbb{R}^d .
$\ \cdot\ _\infty$	supremum-norm of a function.
$B(x, r)$	open ball of radius r centered at x .
$\text{diam}(\mathbb{X})$	diameter of a set $\mathbb{X} \subset \mathbb{R}^d$, defined as $\sup_{x,y \in \mathbb{X}} \ x - y\ $.

- $|\cdot|$ total variation of a measure.
- $\#$ cardinality of a set.
- $\text{Lip}(f)$ Lipschitz constant of a Lipschitz function f .

The rest of the paper is organized as follows. In Section 2, some background on persistent homology is briefly described. The stability results are then discussed in Section 3 whereas the convergence results related to the asymptotic behavior of the sample-based linear representations are stated in Section 4. Section 5 presents some discussion. Proofs can be found in Section 6.

2. BACKGROUND ON PERSISTENCE DIAGRAMS

Persistent homology deals with the evolution of homology through a sequence of topological spaces. We use the field of two elements \mathbb{F}_2 to build the homology groups. A filtration $\mathcal{K} = (K^r)_{r \geq 0}$ is an increasing right-continuous sequence of topological spaces : $K^{r'} \subset K^r$ iff $r' \leq r$ and $K^r = \bigcap_{r' < r} K^{r'}$. For any $q \geq 0$, the inclusion of spaces give rise to linear maps between corresponding homology groups $H_q(K^r)$. The persistence diagram $\text{dgm}_q[\mathcal{K}]$ of the filtration is a succinct way to summarize the evolution of the homology groups. It is a multiset of points in $\mathbb{R}_{>}^2 = \{r = (r_1, r_2) \in \mathbb{R}^2, r_1 < r_2\}^1$, so that each point $r = (r_1, r_2)$ corresponds informally to a q -dimensional "hole" in the filtration \mathcal{K} that appears (or is born) at r_1 and disappears (or dies) at r_2 . The persistence $\text{pers}(r)$ of r is defined as $r_2 - r_1$ and is understood as the lifetime of the corresponding hole. Persistence diagrams are known to exist given mild assumptions on the filtration (see [10, Section 3.8]). Some basic descriptors of persistence diagrams include the α -th total persistence of a diagram, defined as

$$(2.1) \quad \text{Pers}_\alpha(D) := D(\text{pers}^\alpha) = \sum_{r \in D} \text{pers}(r)^\alpha, \quad \alpha > 0,$$

and the persistent Betti numbers, defined as

$$(2.2) \quad \beta^{r,s}(D) := D([0, r] \times (s, \infty)) = \sum_{r \in D} \mathbf{1}\{r \in [0, r] \times (s, \infty)\}, \quad 0 \leq r \leq s.$$

Also, for $M \geq 0$, define

$$(2.3) \quad \text{Pers}_\alpha(D, M) := D(\text{pers}^\alpha \mathbf{1}\{\text{pers} \geq M\}).$$

Given a subset \mathbb{X} of a metric space (\mathbb{Y}, d) , standard constructions of filtrations are the Čech filtration $C(\mathbb{X}) = (C^r(\mathbb{X}))_{r \geq 0}$ and the Rips filtration $\mathcal{R}(\mathbb{X}) = (R^r(\mathbb{X}))_{r \geq 0}$:

$$(2.4) \quad C^r(\mathbb{X}) = \left\{ \text{finite } \sigma \subset \mathbb{X}, \bigcap_{x \in \sigma} B(x, r) \neq \emptyset \right\} \text{ and}$$

$$(2.5) \quad R^r(\mathbb{X}) = \{ \text{finite } \sigma \subset \mathbb{X}, \text{diam}(\sigma) \leq r \},$$

¹Persistence diagrams are in all generality multiset of points in $\{r = (r_1, r_2), -\infty \leq r_1 < r_2 \leq \infty\}$. We only consider diagrams which do not contain points "at infinity" throughout the paper.

where the abstract simplicial complexes on the right are identified with their geometric realizations. The dimension of a simplex σ is equal to $\#\sigma - 1$. If K is a simplicial complex, the set of its simplexes of dimension q is denoted by K_q .

The space of persistence diagrams \mathcal{D} is the set of all finite multisets in $\mathbb{R}_{>}^2$. Wasserstein distances are standard distances on \mathcal{D} . For $p \geq 1$, they are defined as:

$$(2.6) \quad W_p(D, D') := \min_{\gamma} \left(\sum_{r \in D \cup \Delta} \|r - \gamma(r)\|^p \right)^{1/p},$$

where Δ is the diagonal of \mathbb{R}^2 and $\gamma : D \cup \Delta \rightarrow D' \cup \Delta$ is a bijection. The definition is extended to $p = \infty$ by

$$(2.7) \quad W_{\infty}(D, D') = \min_{\gamma} \max_{r \in D \cup \Delta} \|r - \gamma(r)\|,$$

which is called the *bottleneck distance*.

The use of Wasserstein distances is motivated by crucial stability properties they satisfy. Let $f, g : \mathbb{X} \rightarrow \mathbb{R}$ be two continuous functions on a triangulable space \mathbb{X} . Assuming that the persistence diagrams $\text{dgm}_q[f]$ and $\text{dgm}_q[g]$ of the filtrations defined by the sublevel sets of f and g exist and are finite (a condition called *tameness*², the stability property of [15, Main Theorem] asserts that $W_{\infty}(\text{dgm}_q[f], \text{dgm}_q[g]) \leq \|f - g\|_{\infty}$, i.e. the diagrams are stable with respect to the functions they are built with. The functions f and g have to be thought of as representing the data: for instance, if the Čech filtration is built on a data set $\mathbb{X}_n = \{X_1, \dots, X_n\}$, then $\text{dgm}_q[C(\mathbb{X}_n)] = \text{dgm}_q[f]$ where f is the distance function to \mathbb{X}_n , i.e. $f(\cdot) = d(\cdot, \mathbb{X}_n)$. When $p < \infty$, similar stability results have been proved under more restrictive conditions on the ambient space \mathbb{X} , which we now detail.

Definition 2.1. *A metric space \mathbb{X} is said to have bounded m -th total persistence if there exists a constant $C_{\mathbb{X},m}$ such that for all tame 1-Lipschitz functions $f : \mathbb{X} \rightarrow \mathbb{R}$, and, for all $q \geq 1$, $\text{Pers}_m(\text{dgm}_q[f]) \leq C_{\mathbb{X},m}$.*

This assumption holds, for instance, for a d -dimensional manifold \mathbb{X} when $m > d$ with

$$(2.8) \quad C_{\mathbb{X},m} = \frac{m}{m-d} C_{\mathbb{X}} \text{diam}(\mathbb{X})^{m-d},$$

$C_{\mathbb{X}}$ being a constant depending only on \mathbb{X} (see [16]). The stability theorem for the p -th Wasserstein distances claims:

Theorem 2.2 (Section 3 of [16]). *Let \mathbb{X} be a compact triangulable metric space with bounded m -th total persistence for some $m \geq 1$. Let $f, g : \mathbb{X} \rightarrow \mathbb{R}$ be two tame Lipschitz functions. Then, for $q \geq 0$,*

$$(2.9) \quad W_p(\text{dgm}_q[f], \text{dgm}_q[g]) \leq C_0^{\frac{1}{p}} \|f - g\|_{\infty}^{1 - \frac{m}{p}},$$

for $p \geq m$, where $C_0 = C_{\mathbb{X},m} \max\{\text{Lip}(f)^m, \text{Lip}(g)^m\}$.

²Tameness holds under simple conditions, see [10, Section 3.9], which we will always assume to hold in the following)

3. STABILITY RESULTS FOR LINEAR REPRESENTATIONS

In [25, Corollary 12], representations of diagrams are shown to be Lipschitz with respect to the 1 Wasserstein distance for weight functions of the form $w(r) = \arctan(B \cdot \text{pers}(r)^\alpha)$ with $\alpha > m + 1, B > 0$, provided the diagrams are built with the sublevels of functions defined on a space \mathbb{X} having bounded m -th total persistence. The stability result is proved for a particular function $\phi : \mathbb{R}_>^2 \rightarrow \mathcal{B}$ defined by $\phi(r) = K(r, \cdot)$, with K a bounded Lipschitz kernel and \mathcal{B} the associated RKHS (short for Reproducing Kernel Hilbert Space, see [2] for a monograph on the subject). We present a generalization of the stability result to (i) general weight functions w , (ii) any bounded Lipschitz function ϕ , and (iii) we only require $\alpha > m$.

Consider weight functions $w : \mathbb{R}_>^2 \rightarrow \mathbb{R}_+$ of the form $w(r) = \tilde{w}(\text{pers}(r))$ for a differentiable function $\tilde{w} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfying $\tilde{w}(0) = 0$, and, for some $A > 0, \alpha \geq 1$,

$$(3.1) \quad \forall u \geq 0, |\tilde{w}'(u)| \leq Au^{\alpha-1}.$$

Examples of such functions include $w(r) = \arctan(B \cdot \text{pers}(r)^\alpha)$ for $B > 0$ and $w(r) = \text{pers}(r)^\alpha$. We denote the class of such weight functions by $\mathbb{W}(\alpha, A)$. In contrast to [25], the function ϕ does not necessarily take its values in a RKHS, but simply in a Banach space (so that its Bochner integral –see for instance [18, Chapter 4]– is well defined).

Theorem 3.1. *Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a Banach space, and let $\phi : \mathbb{R}_>^2 \rightarrow \mathcal{B}$ be a Lipschitz continuous function. Furthermore, for $w \in \mathbb{W}(\alpha, A)$ with $A > 0, \alpha \geq 1$, let $\Phi_w(D) := D(w\phi)$, and for two persistence diagrams D_1 and D_2 let $G\{t\} := \max\{\text{Pers}_t(D_1), \text{Pers}_t(D_2)\}, t \geq 0$. Then, for $1 \leq p \leq \infty$, and $a \in [0, 1]$ (and using the conventions $0/\infty = 0$ and $\infty/\infty = 1$), we have*

$$(3.2) \quad \|\Phi_w(D_1) - \Phi_w(D_2)\|_{\mathcal{B}} \leq \text{Lip}(\phi) \frac{A}{\alpha} \left(G \left\{ p \frac{\alpha}{p-1} \right\} \right)^{1-\frac{1}{p}} W_p(D_1, D_2) \\ + \|\phi\|_{\infty} A \left(2G \left\{ p \frac{\alpha-a}{p-a} \right\} \right)^{1-\frac{a}{p}} W_p(D_1, D_2)^a.$$

The quantity $G\{q\}$ can often be controlled. For instance, if the diagrams are built with Lipschitz continuous functions $f : \mathbb{X} \rightarrow \mathbb{R}$, and \mathbb{X} is a space having bounded m -th total persistence.

Corollary 3.2. *Let $A > 0, \alpha \geq 1, q \geq 0$ and consider a compact triangulable metric space \mathbb{X} having bounded m -th total persistence for some $m \geq 1$. Suppose that $f, g : \mathbb{X} \rightarrow \mathbb{R}$ are two tame Lipschitz continuous functions, $w \in \mathbb{W}(\alpha, A)$, and $a \in [0, 1]$. Then, for $m \leq p \leq \infty$ such that $\alpha \geq m + a \left(1 - \frac{m}{p}\right) \geq 0$, if $C_0 = C_{\mathbb{X}, m} \max\{\text{Lip}(f)^m, \text{Lip}(g)^m\}$ and ℓ is the maximum persistence in the two diagrams $\text{dgm}_q[f], \text{dgm}_q[g]$:*

$$(3.3) \quad \|\Phi_w(\text{dgm}_q[f]) - \Phi_w(\text{dgm}_q[g])\|_{\mathcal{B}} \leq C_1 W_p(\text{dgm}_q[f], \text{dgm}_q[g]) \\ + C_2 W_p(\text{dgm}_q[f], \text{dgm}_q[g])^a,$$

where $C_1 = \text{Lip}(\phi) \frac{A}{\alpha} \ell^{\alpha-m(1-\frac{1}{p})} C_0^{1-\frac{1}{p}}$ and $C_2 = \|\phi\|_\infty A \ell^{\alpha-m-a(1-\frac{m}{p})} (2C_0)^{1-\frac{a}{p}}$.

If $\alpha > m + 1$ and $p = \infty$, then the result is similar to Theorem 3.3 in [25]. However, Corollary 3.2 implies that the representations are still continuous (actually Hölder continuous) when $\alpha \in (m, m + 1]$, and this is the novelty of the result. Indeed, for such an α , one can always chose a small enough so that the stability result (3.3) holds. The proofs of Theorem 3.1 and Corollary 3.2 consist of adaptations of similar proofs in [25]. They can be found in Section 6.

Remark 3.3. (a) One cannot expect to obtain an inequality of the form (3.2) without quantities $G\{t\}$ (or other quantities depending on the diagrams) appearing on the right-hand side. For instance, in the case $p = \infty$, it is clear that adding an arbitrary number of points near the diagonal will not change the bottleneck distance between the diagram, whereas the distance between representations can become arbitrarily large.

(b) Laws of large numbers stated in the next section (see also Theorem 1.1 already stated in the introduction), show that Theorem 3.1 is optimal: take $w = \text{pers}^\alpha$ and $\phi \equiv 1$. If \mathbb{X}_n is a sample on the d -dimensional cube $[0, 1]^d$ (which has bounded m -th total persistence for $m > d$), then $\Phi_w(\text{dgm}_q[C(\mathbb{X}_n)]) = \text{Pers}_\alpha(\text{dgm}_q[C(\mathbb{X}_n)])$. The quantity $\text{Pers}_\alpha(\text{dgm}_q[C(\mathbb{X}_n)])$ does not converge to 0 for $\alpha \leq d$ (it even diverges if $\alpha < d$), whereas the bottleneck distance between $\text{dgm}_q[C(\mathbb{X}_n)]$ and the empty diagram does converge to 0.

The following corollary to the stability result also is a contribution to the asymptotic study of next section. It presents rates of convergence of representations in a random setting. Let $\mathbb{X}_n = \{X_1, \dots, X_n\}$ be a n -sample of i.i.d. points from a distribution on some manifold \mathbb{X} . We are interested in the convergence of representations $\Phi_w(\text{dgm}_q[C(\mathbb{X}_n)])$ to the representations $\Phi_w(\text{dgm}_q[C(\mathbb{X})])$. The nerve theorem asserts that for any subspace $\mathbb{X}' \subset \mathbb{X}$, $\text{dgm}_q[C(\mathbb{X}')] = \text{dgm}_q[f]$ where $f(x)$ is the distance from $x \in \mathbb{X}$ to \mathbb{X}' . We obtain the following corollary, whose proof is found in Section 6:

Corollary 3.4. *Consider a d -dimensional compact Riemannian manifold \mathbb{X} , and let $\mathbb{X}_n = \{X_1, \dots, X_n\}$ be a n -sample of i.i.d. points from a distribution having a density κ with respect to the d -dimensional Hausdorff measure on \mathbb{X} . Assume that $0 < \inf \kappa \leq \sup \kappa < \infty$. Let $w \in \mathbb{W}(\alpha, A)$ for some $A > 0, \alpha > d$, and let $\phi : \mathbb{R}_>^2 \rightarrow \mathcal{B}$ be a Lipschitz function. Then, for $q \geq 0$, and for n large enough,*

$$(3.4) \quad E \left[\|\Phi_w(\text{dgm}_q[C(\mathbb{X}_n)]) - \Phi_w(\text{dgm}_q[C(\mathbb{X})])\|_{\mathcal{B}} \right] \leq C \|\phi\|_\infty \frac{\alpha}{\alpha - d} \left(\frac{\ln n}{n} \right)^{\frac{\alpha}{d}-1},$$

where C is a constant depending on \mathbb{X} , A and the density κ .

4. CONVERGENCE OF TOTAL PERSISTENCE

Consider again the i.i.d. model: let $\mathbb{X}_n = \{X_1, \dots, X_n\}$ be i.i.d. observations of density κ with respect to the d -dimensional Hausdorff measure on some d -dimensional manifold \mathbb{X} . The general question we are addressing in this section

is the convergence of the observed diagrams $\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]$ to $\text{dgm}_q[\mathcal{K}(\mathbb{X})]$, with \mathcal{K} either the Rips or the Čech filtration. Of course, the question has already been answered in some sense. For instance, Theorem 2.2 affirms that the sequence of observed diagrams will always converge to $\text{dgm}_q[\mathcal{K}(\mathbb{X})]$ for the bottleneck distance, if \mathcal{K} is the Čech filtration³. However, this is not informative with respect to the convergence of the representations introduced in the previous section, which is related to a weak convergence of measure: For which functions ϕ does $\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)](\phi)$ converge to $\text{dgm}_q[\mathcal{K}(\mathbb{X})](\phi)$?

The stability theorem for the bottleneck distance asserts that, for small $\varepsilon > 0$, and for n large enough, $\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]$ can be decomposed into two separate sets of points: a set of fixed size $\text{dgm}_{\text{true},q}[\mathcal{K}(\mathbb{X}_n)]$ that is ε -close to points in $\text{dgm}_q[\mathcal{K}(\mathbb{X})]$ and the remaining part of the diagram, $\text{dgm}_{\text{noise},q}[\mathcal{K}(\mathbb{X}_n)]$, usually consisting of a large number of points, which have persistence smaller than ε , i.e. these are the points that lie close to the diagonal. A Taylor expansion of ϕ shows that the difference between $\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)](\phi)$ and $\text{dgm}_q[\mathcal{K}(\mathbb{X})](\phi)$ is of the order of $\text{dgm}_{\text{noise},q}[\mathcal{K}(\mathbb{X}_n)](\text{pers}^\alpha)$ for some $\alpha \geq 0$. The latter quantities are therefore of utmost interest to achieve our goal. Instead of directly studying $\text{dgm}_{\text{noise},q}[\mathcal{K}(\mathbb{X}_n)](\text{pers}^\alpha)$ for \mathbb{X}_n on a d -dimensional manifold, we focus on the study of the quantity $\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)](\text{pers}^\alpha)$ for \mathbb{X}_n in a cube $[0, 1]^d$.

Contributions to the study of quantities of the form $\text{dgm}_q[\mathcal{K}(\mathbb{S}_n)](\phi)$ have been made in [22], where \mathbb{S}_n is considered to be the restriction of a stationary process to a box of volume n in \mathbb{R}^d . Specifically, [22] shows the vague convergence of the rescaled diagram $n^{-1} \text{dgm}_q[\mathcal{K}(\mathbb{S}_n)]$ to some Radon measure μ_q . The two recent papers [21, 35] prove that a similar convergence actually holds for \mathbb{S}_n a binomial sample on a manifold. However, vague convergence deals with continuous functions ϕ with compact support, whereas we are interested in functions of the type pers^α , which are not even bounded. Our contributions to the matter consists in proving, for samples on the cube $[0, 1]^d$, a stronger convergence, allowing test functions to have non-compact support and polynomial growth. As a gentle introduction to the formalism used later, we first recall some known results from geometric probability on the study of Betti numbers, and we also detail relevant results of [21, 22, 35].

4.1. Prior work. In the following, \mathcal{K} refers to either the Čech or the Rips filtration. Let κ be a density on $[0, 1]^d$ such that:

$$(4.1) \quad 0 < \inf \kappa \leq \sup \kappa < \infty.$$

Note that the cube $[0, 1]^d$ could be replaced by any compact convex body (i.e. the boundary of an open bounded convex set). However, the proofs (especially geometric arguments of Section 6.4) become much more involved in this greater

³A similar result states that the bottleneck distance between two diagrams, each built with the Rips filtration on some space, is controlled by the Hausdorff distance between the two spaces (see Theorem 3.1 [9]). As the Rips filtration, contrary to the Čech filtration, cannot be seen as the filtration of the sublevel sets of some function, this stability is not a consequence of Theorem 2.2.

generality. To keep the main ideas clear, we therefore restrict ourselves to the case of the cube. We indicate, however, when challenges arise in the more general setting.

Let $(X_i)_{i \geq 1}$ be a sequence of i.i.d. random variables sampled from density κ and let $(N_i)_{i \geq 1}$ be an independent sequence of Poisson variables with parameter i . In the following \mathbb{X}_n denotes either $\{X_1, \dots, X_n\}$, a binomial process of intensity κ and of size n , or $\{X_1, \dots, X_{N_n}\}$, a Poisson process of intensity $n\kappa$. The fact that the binomial and Poisson processes are built in this fashion is not important for weak laws of large numbers (only the law of the variables is of interest), but it is crucial for strong laws of large numbers to make sense.

The persistent Betti numbers $\beta_q^{r,s}(\text{dgm}_q[\mathcal{K}]) = \text{dgm}_q[\mathcal{K}(\mathbb{X}_n)](\mathbf{1}_{[r,\infty) \times [s,\infty)})$ are denoted more succinctly by $\beta_q^{r,s}(\mathcal{K})$. When $r = s$, we use the notation $\beta_q^r(\mathcal{K})$.

Theorem 4.1 (Theorem 1.4 in [35]). *Let $r > 0$ and $q \geq 0$. Then, with probability one, $n^{-1} \beta_q^r(\mathcal{K}(n^{1/d}\mathbb{X}_n))$ converges to some constant. The convergence also holds in expectation.*

The theorem is originally stated with the Čech filtration but its generalization to the Rips filtration (or even to more general filtrations considered in [22]) is straightforward. The proof of this theorem is based on a simple, yet useful geometric lemma, which still holds for the persistent Betti numbers, as proven in [22]. Recall that for $j \geq 0$, K_j denotes the j -skeleton of the simplicial complex K .

Lemma 4.2 (Lemma 2.11 in [22]). *Let $\mathbb{X} \subset \mathbb{Y}$ be two subsets of \mathbb{R}^d . Then*

$$(4.2) \quad |\beta_q^{r,s}(\mathcal{K}(\mathbb{X})) - \beta_q^{r,s}(\mathcal{K}(\mathbb{Y}))| \leq \sum_{j=q}^{q+1} |K_j^s(\mathbb{Y}) \setminus K_j^s(\mathbb{X})|.$$

In [22], this lemma was used to prove the convergence of expectations of diagrams of stationary point processes. As indicated in [21, Remark 2.4], this lemma can also be used to prove the convergence of the expectations of diagrams for non-homogeneous binomial processes on manifold. Let $C_c(\mathbb{R}_>^2)$ be the set of functions $\phi : \mathbb{R}_>^2 \rightarrow \mathbb{R}$ with compact support. We say that a sequence $(\mu_n)_{n \geq 0}$ of measures on $\mathbb{R}_>^2$ converges C_c -vaguely to μ if $\forall \phi \in C_c(\mathbb{R}_>^2)$, $\mu_n(\phi) \xrightarrow{n \rightarrow \infty} \mu(\phi)$. Note that this does not include the function $\phi = 1$ or the function $\phi = \text{pers}$. Vague convergence is denoted by $\xrightarrow{v_c}$. Set $\mu_n = n^{-1} \text{dgm}[\mathcal{K}(n^{1/d}\mathbb{X}_n)]$. Remark 2.4 in [21] implies the following theorem.

Theorem 4.3 (Remark 2.4 in [21] and Theorem 1.5 in [22]). *Let κ be a probability density function on a d -dimensional compact C^1 manifold \mathbb{X} , with $\int_{\mathbb{X}} \kappa^j(z) dz < \infty$ for $j \in \mathbb{N}$. Then, for $q \geq 0$, there exists a unique Radon measure μ_q^κ on $\mathbb{R}_>^2$ such that*

$$(4.3) \quad E[\mu_n] \xrightarrow[n \rightarrow \infty]{v_c} \mu_q^\kappa$$

and

$$(4.4) \quad \mu_n \xrightarrow[n \rightarrow \infty]{v_c} \mu_q^\kappa \text{ a.s..}$$

The measure μ_q^κ is called the persistence diagram of intensity κ for the filtration \mathcal{K} .

4.2. Main results. A function $\phi : \mathbb{R}_>^2 \rightarrow \mathbb{R}$ is said to vanish on the diagonal if

$$(4.5) \quad \lim_{\varepsilon \rightarrow 0} \sup_{\text{pers}(r) \leq \varepsilon} |\phi(r)| = 0.$$

Denote by $C_0(\mathbb{R}_>^2)$ the set of all such functions. The weight functions of Section 3 all lie in $C_0(\mathbb{R}_>^2)$. We say that a function $\phi : \mathbb{R}_>^2 \rightarrow \mathbb{R}$ has polynomial growth if there exist two constants $A, \alpha > 0$, such that

$$(4.6) \quad |\phi(r)| \leq A (1 + \text{pers}(r)^\alpha).$$

The class $C_{\text{poly}}(\mathbb{R}_>^2)$ of functions in $C_0(\mathbb{R}_>^2)$ with polynomial growth constitutes a reasonable class of functions $w \cdot \phi$ one may want to build a representation with. Our goal is to extend the convergence of Theorem 4.3 to this larger class of functions. Convergence of measures μ_n to μ with respect to $C_{\text{poly}}(\mathbb{R}_>^2)$, i.e. $\forall \phi \in C_{\text{poly}}(\mathbb{R}_>^2), \mu_n(\phi) \xrightarrow[n \rightarrow \infty]{v_p} \mu(\phi)$, is denoted by $\xrightarrow{v_p}$. Note that this class of functions is standard: it is for instance known to characterize p -th Wasserstein convergence in optimal transport (see [36, Theorem 6.9]).

Theorem 4.4. (i) For $q \geq 0$, there exists a unique Radon measure μ_q^κ such that $E[\mu_q^n] \xrightarrow[n \rightarrow \infty]{v_p} \mu_q^\kappa$ and, with probability one, $\mu_q^n \xrightarrow[n \rightarrow \infty]{v_p} \mu_q^\kappa$. The measure μ_q^κ is called the q -th persistence diagram of intensity κ for the filtration \mathcal{K} . It does not depend on whether \mathbb{X}_n is a Poisson or a binomial process, and is of positive finite mass.

(ii) The convergence also holds pointwise for the L_p distance: for all $\phi \in C_{\text{poly}}(\mathbb{R}_>^2)$, and for all $p \geq 1$, $\mu_q^n(\phi) \xrightarrow[n \rightarrow \infty]{L_p} \mu_q^\kappa(\phi)$. In particular, $|\mu_q^\kappa(\phi)| < \infty$.

Remark 4.5. (a) Remark 2.4 together with Theorem 1.1 in [21] imply that the measure μ_q^κ has the following expression:

$$(4.7) \quad \mu_q^\kappa(\phi) = E \left[\int_{\mathbb{R}_>^2} \phi(r\kappa(X)^{-1/d}) d\mu_q(r) \right] \quad \forall \phi \in C_c(\mathbb{R}_>^2),$$

where $\mu_q = \mu_q^1$ is the q -th persistence diagram of uniform density on $[0, 1]^d$, appearing in Theorem 4.3, and the expectation is taken with respect to a random variable X having a density κ .

(b) Assume $q = 0$ and $d = 1$. Then, the persistence diagram $\text{dgm}_0[\mathcal{K}(\mathbb{X}_n)]$ is simply the collection of the intervals $(X_{(i+1)} - X_{(i)})$ where $X_{(1)} < \dots < X_{(n)}$ is the order statistics of \mathbb{X}_n . The measure $E[\mu_0^n]$ can be explicitly computed: it converges to a measure having density $u \mapsto E[\exp(-u\kappa(X))\kappa(X)]$ with respect to the Lebesgue measure on \mathbb{R}_+ , where X has density κ . Take κ the uniform

density on $[0, 1]$: one sees that this is coherent with the basic fact that the spacings of a homogeneous Poisson process on \mathbb{R} are distributed according to an exponential distribution. Moreover, the expression (4.7) is found again in this special case.

(c) Theorem 1.9 in [22] states that the support of μ_q^1 is $\mathbb{R}_{>}^2$. Using equation (4.7), the same holds for μ_q^κ .

(d) Theorem 1.1 is a direct corollary of Theorem 4.4. Indeed, we have

$$\begin{aligned} n^{\frac{\alpha}{d}-1} \text{Pers}_\alpha(\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]) &:= n^{\frac{\alpha}{d}-1} \sum_{r \in \text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]} \text{pers}^\alpha(r) \\ &= n^{-1} \sum_{r \in \text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]} \text{pers}^\alpha(n^{-\frac{1}{d}}r) \\ &= \mu_q^n(\text{pers}^\alpha), \end{aligned}$$

a quantity which converges to $\mu_q^\kappa(\text{pers}^\alpha)$. The relevance of Theorem 1.1 is illustrated in Figure 2, where Čech complexes are computed on random samples on the torus.

The core of the proof of Theorem 4.4 consists in a control of the number of points appearing in diagrams. This bound is obtained thanks to geometric properties satisfied by the Čech and Rips filtrations. Finding good requirements to impose on a filtration \mathcal{K} for this control to hold is an interesting question. The following states some non-asymptotic controls of the number of points in diagrams which are interesting by themselves.

Proposition 4.6. *Let $M \geq 0$ and define $U_M = \mathbb{R} \times [M, \infty)$. Then, there exists constants $c_1, c_2 > 0$ (which can be made explicit) depending on κ and q , such that, for any $t > 0$,*

$$(4.8) \quad P(\mu_q^n(U_M) > t) \leq c_1 \exp(-c_2(M^d + t^{1/(q+1)})).$$

As an immediate corollary, the moments of the total mass $|\mu_n|$ are uniformly bounded. However, the proof of the almost sure finiteness of $\sup_n |\mu_n|$ is much more intricate. Indeed, we are unable to control directly this quantity, and we prove that a majorant of $|\mu_n|$ satisfies concentration inequalities. The majorant arises as the number of simplicial complexes of a simpler process, whose expectation is also controlled.

It is natural to wonder whether μ_q^κ has some density with respect to the Lebesgue measure on $\mathbb{R}_{>}^2$: it is the case for the for $d = 1$, and it is shown in [11] that $E[\mu_q^n]$ also has a density. Even if those elements are promising, it is not clear whether the limit μ_q^κ has a density in a general setting. However, we are able to prove that the marginals of μ_q^κ have densities.

Proposition 4.7. *Let π_1 (resp. π_2) be the projection on the x -axis (resp. y -axis). Then, for $q > 0$, the pushforwards $\pi_1^\star(\mu_q^\kappa)$ and $\pi_2^\star(\mu_q^\kappa)$ have densities with respect to the Lebesgue measure on \mathbb{R} . For $q = 0$, $\pi_2^\star(\mu_q^\kappa)$ has a density.*

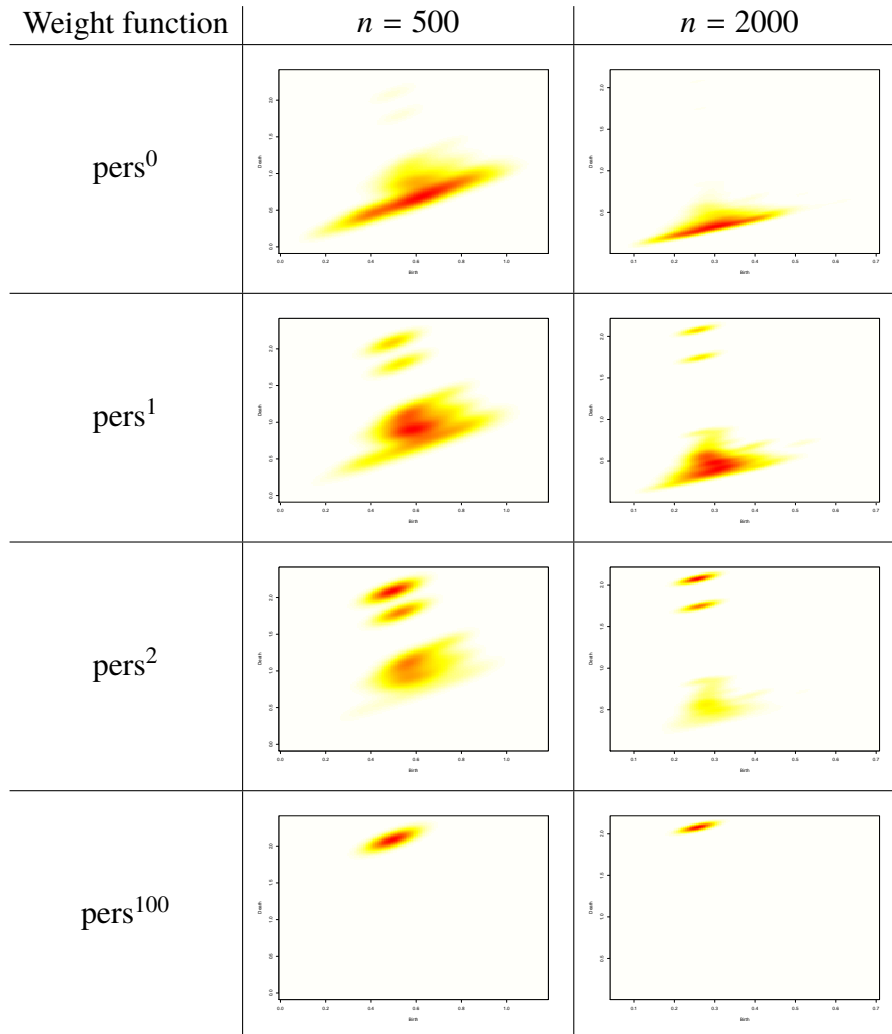


FIGURE 2. For $n = 500$ or 2000 points uniformly sampled on the torus, persistence images [1] for different weight functions are displayed. For $\alpha < 2$, the mass of the topological noise is far larger than the mass of the true signal, the latter being comprised by the two points with high-persistence. For $\alpha = 2$, the two points with high-persistence are clearly distinguishable. For $\alpha = 100$, the noise has also disappeared, but so has one of the point with high-persistence.

5. DISCUSSION

The tuning of the weight functions in the representations of persistence diagrams is a critical issue in practice. When the statistician has good reasons to believe that the data lies near a d -dimensional structure, we give, through two different approaches, an heuristic to tune this weight function: a weight of the form pers^α with $\alpha \geq d$ is sensible. The study carried out in this paper allowed us to show new results on the asymptotic structure of random persistence diagrams.

While the existence of a limiting measure in a weak sense was already known, we strengthen the convergence, allowing a much larger class of test functions. Some results about the properties of the limit are also shown, namely that it has a finite mass, finite moments, and that its marginals have densities with respect to the Lebesgue measure. Challenging open questions include:

- Convergence of the rescaled diagrams μ_n with respect to some transport metric: The main issue consists in showing that one can extend, in a meaningful way, the distance W_p to general Radon measures. This is the topic of a recent work (see Section 5.1 in [19]).
- Existence of a density for the limiting measure: An approach for obtaining such results would be to control the numbers of points of a diagram in some square $[r_1, r_2] \times [s_1, s_2]$.
- Convergence of the number of points in the diagrams: The number of points in the diagrams is a quantity known to be not stable (motivating the use of bottleneck distances, which is blind to them). However, experiments show that this number, conveniently rescaled, converges in this setting. An analog of Lemma 4.2 for the number of points in the diagrams with small persistence would be crucial to attack this problem.
- Generalization to manifolds: While the vague convergence of the rescaled diagrams is already proven in [21], allowing test functions without compact support seems to be a challenge. Once again, the crucial issue consists in controlling the total number of points in the diagrams.
- Dimension estimation: We have proved that the total persistence of a diagram built on a given point cloud depends crucially on the intrinsic dimension of such a point cloud. Inferring the dependence of the total persistence with respect to the size of the point cloud (through subsampling) leads to estimators of this intrinsic dimension. Studying the properties of such estimators is the topic of an on-going work of Henry Adams and co-authors (personal communication).

6. PROOFS

6.1. Proof of Theorem 3.1. We only treat the case $p < \infty$, the proof being easily adapted to the case $p = \infty$. Introduce for μ, ν two measures of mass $m > 0$ on $\mathbb{R}_>^2$, the Monge-Kantorovitch distance between μ and ν :

$$(6.1) \quad d_{\text{MK}}(\mu, \nu) := \sup \left\{ \mu(\phi) - \nu(\phi), \phi : \mathbb{R}_>^2 \rightarrow \mathbb{R} \text{ 1-Lipschitz} \right\}.$$

Fix two persistence diagrams D_1 and D_2 . Denote $\mu = D_1(w \cdot)$ (resp. $\nu = D_2(w \cdot)$) the measure having density w with respect to D_1 (resp. D_2). For γ a matching attaining the p -th Wasserstein distance between D_1 and D_2 , denote $\tilde{\mu} = \sum_{r \in D_1 \cup \Delta} w(\gamma(r)) \delta_r$. We have

$$(6.2) \quad \begin{aligned} \|\Phi_w(D_1) - \Phi_w(D_2)\|_{\mathcal{B}} &= \|\mu(\phi) - \nu(\phi)\|_{\mathcal{B}} \\ &\leq \|\mu(\phi) - \tilde{\mu}(\phi)\|_{\mathcal{B}} + \|\tilde{\mu}(\phi) - \nu(\phi)\|_{\mathcal{B}} \\ &\leq \|\phi\|_{\infty} |\mu - \tilde{\mu}| + \text{Lip}(\phi) d_{\text{MK}}(\tilde{\mu}, \nu). \end{aligned}$$

We bound the two terms in the sum separately. Let us first bound $d_{\text{MK}}(\tilde{\mu}, \nu)$. The Monge-Kantorovitch distance is also the infimum of the costs of transport plans between $\tilde{\mu}$ and ν (see [36, Chapter 2] for details), so that

$$d_{\text{MK}}(\tilde{\mu}, \nu) \leq \sum_{r \in D_1 \cup \Delta} w(\gamma(r)) \|r - \gamma(r)\|.$$

Define q such that $\frac{1}{p} + \frac{1}{q} = 1$. As condition (3.1) implies that $\|w(r)\| \leq \frac{A}{\alpha} \text{pers}(r)^\alpha$, the distance $d_{\text{MK}}(\tilde{\mu}, \nu)$ is bounded by

$$\begin{aligned} \sum_{r \in D_1 \cup \Delta} w(\gamma(r)) \|r - \gamma(r)\| &\leq \left(\sum_{r \in D_1 \cup \Delta} w(\gamma(r))^q \right)^{1/q} \left(\sum_{r \in D_1 \cup \Delta} \|r - \gamma(r)\|^p \right)^{1/p} \\ &\leq \frac{A}{\alpha} \left(\sum_{r \in D_1 \cup \Delta} \text{pers}(\gamma(r))^{q\alpha} \right)^{1/q} W_p(D_1, D_2) \\ (6.3) \quad &\leq \frac{A}{\alpha} (G\{q\alpha\})^{1/q} W_p(D_1, D_2). \end{aligned}$$

We now treat the first part of the sum in (6.2). For r_1, r_2 , in $\mathbb{R}_>^2$ with $\text{pers}(r_1) \leq \text{pers}(r_2)$, define the path with unit speed $h : [\text{pers}(r_1), \text{pers}(r_2)] \rightarrow \mathbb{R}_>^2$ by

$$h(t) = r_2 \frac{t - \text{pers}(r_1)}{\text{pers}(r_2) - \text{pers}(r_1)} + r_1 \frac{\text{pers}(r_2) - t}{\text{pers}(r_2) - \text{pers}(r_1)},$$

so that it satisfies $\text{pers}(h(t)) = t$. The quantity $|w(r_1) - w(r_2)|$ is bounded by

$$\begin{aligned} \int_{\text{pers}(r_1)}^{\text{pers}(r_2)} |\nabla w(h(t)) \cdot h'(t)| dt &\leq \int_{\text{pers}(r_1)}^{\text{pers}(r_2)} A \text{pers}(h(t))^{\alpha-1} dt \\ &\leq \int_{\text{pers}(r_1)}^{\text{pers}(r_2)} A t^{\alpha-1} dt \\ &= \frac{A}{\alpha} (\text{pers}(r_2)^\alpha - \text{pers}(r_1)^\alpha). \end{aligned}$$

For $0 < y < x$ and $0 \leq a \leq 1$, using the convexity of $t \mapsto t^\alpha$, it is easy to see that $x^\alpha - y^\alpha \leq \alpha(x - y)^a x^{\alpha-a}$. Define $p' = \frac{p}{a}$, $q' = \frac{p'}{p'-1}$ and $M(r) := \max(\text{pers}(r), \text{pers}(\gamma(r)))$. We have,

$$\begin{aligned} |\mu - \tilde{\mu}| &= \sum_{r \in D_1 \cup \Delta} |w(r) - w(\gamma(r))| \\ &\leq A \sum_{r \in D_1 \cup \Delta} |\text{pers}(r) - \text{pers}(\gamma(r))|^a M(r)^{\alpha-a} \\ &\leq A \left(\sum_{r \in D_1 \cup \Delta} |\text{pers}(r) - \text{pers}(\gamma(r))|^{ap'} \right)^{1/p'} \left(\sum_{r \in D_1 \cup \Delta} M(r)^{q'(\alpha-a)} \right)^{1/q'} \\ &\leq A W_{ap'}(D_1, D_2)^a \left(\sum_{r \in D_1 \cup \Delta} (\text{pers}(r)^{q'(\alpha-a)} + \text{pers}(\gamma(r))^{q'(\alpha-a)}) \right)^{1/q'} \end{aligned}$$

$$(6.4) \quad \leq AW_{ap'}(D_1, D_2)^a 2^{1/q'} (G\{q'(\alpha - a)\})^{1/q'}$$

Combining equations (6.2), (6.3) and (6.4) concludes the proof.

6.2. Proof of Corollary 3.2. Corollary 3.2 follows easily by using the definition of a space with bounded m -th total persistence along with the inequality $G\{t_1 + t_2\} \leq \ell^1 G\{t_2\}$.

6.3. Proof of Corollary 3.4. As already discussed, Theorem 3.1 can be applied with $f_n = d(\cdot, \mathbb{X}_n)$ and f the null function on the manifold \mathbb{X} . Take $p = \infty$, $d < \alpha$ and $0 < a < \min(1, \alpha - d)$:

$$(6.5) \quad \begin{aligned} & \|\Phi_w(\text{dgm}_q[f_n]) - \Phi_w(\text{dgm}_q[f])\|_{\mathcal{B}} \\ & \leq \text{Lip}(\phi) \frac{A}{\alpha} G\{\alpha\} W_{\infty}(\text{dgm}_q[f_n], \text{dgm}_q[f]) \\ & \quad + 2\|\phi\|_{\infty} A G\{\alpha - a\} W_{\infty}(\text{dgm}_q[f_n], \text{dgm}_q[f])^a. \end{aligned}$$

It is mentioned at the end of Section 2 in [16] that, for $m > d$,

$$\text{Pers}_m(\text{dgm}_q[f_n]) \leq m C_{\mathbb{X}} \|f_n\|_{\infty}^{m-d} / (m - d)$$

for some constant $C_{\mathbb{X}}$ depending only on \mathbb{X} . Moreover, the stability theorem for the bottleneck distance ensures that $W_{\infty}(\text{dgm}_q[f_n], \text{dgm}_q[f]) \leq \|f_n\|_{\infty}$. Therefore,

$$(6.6) \quad \begin{aligned} \|\Phi_w(\text{dgm}_q[f_n]) - \Phi_w(\text{dgm}_q[f])\|_{\mathcal{B}} & \leq \text{Lip}(\phi) \frac{\alpha A C_{\mathbb{X}}}{\alpha - d} \|f_n\|_{\infty}^{\alpha-d+1} \\ & \quad + \|\phi\|_{\infty} \frac{2 A C_{\mathbb{X}} (\alpha - a)}{\alpha - a - d} \|f_n\|_{\infty}^{\alpha-a-d+a} \\ & \leq \text{Lip}(\phi) \frac{\alpha A C_{\mathbb{X}}}{\alpha - d} \|f_n\|_{\infty}^{\alpha-d+1} \\ & \quad + \|\phi\|_{\infty} \frac{2 A C_{\mathbb{X}} \alpha}{\alpha - d} \|f_n\|_{\infty}^{\alpha-d}, \end{aligned}$$

where, in the last line, the second term was minimized over a . The quantity $\|f_n\|_{\infty}$ is the Hausdorff distance between \mathbb{X}_n and \mathbb{X} . Elementary techniques of geometric probability (see for instance [17]) show that if \mathbb{X} is a compact d -dimensional manifold, then $E[\|f_n\|_{\infty}^{\beta}] \leq c \left(\frac{\ln n}{n}\right)^{\beta/d}$ for $\beta \geq 0$, where c is some constant depending on β , \mathbb{X} , $\inf \kappa$ and $\sup \kappa$. Therefore, the first term of the sum (6.6) being negligible,

$$\begin{aligned} E \left[\|\Phi_w(\text{dgm}_q[f_n]) - \Phi_w(\text{dgm}_q[f])\|_{\mathcal{B}} \right] & \leq \|\phi\|_{\infty} \frac{2 A C_{\mathbb{X}} \alpha}{\alpha - d} c \left(\frac{\ln n}{n}\right)^{(\alpha-d)/d} \\ & \quad + o \left(\left(\frac{\ln n}{n}\right)^{(\alpha-d)/d} \right). \end{aligned}$$

In particular, the conclusion holds for any $C > 2 A C_{\mathbb{X}} c$, for n large enough. \square

. We now prove the propositions of Section 4. In the following proofs, c is a constant, depending on κ , d and q , which can change from line to line (or even represent two different constants in the same line). A careful read can make all those constants explicit. If a constant depends also on some additional parameter x , it is then denoted by $c(x)$.

6.4. Proof of Proposition 4.6. First, as the right hand side of the inequality (4.8) does not depend on n , one may safely assume that μ_n is built with the binomial process. The proof is based on two observations.

(i) Let $r(\sigma) := \min\{r > 0, \sigma \in K^r(\mathbb{X}_n)\}$ denote the *filtration time* of σ . A simplex σ is said to be *negative* in the filtration $\mathcal{K}(\mathbb{X}_n)$ if σ is not included in any cycle of $K^{r(\sigma)}(\mathbb{X}_n)$. A basic result of persistent homology states that points in $\text{dgm}_q[\mathcal{K}(\mathbb{X}_n)]$ are in bijection with pair of simplexes, one negative and one positive (i.e. non-negative). Moreover, the death time r_2 of a point $r = (r_1, r_2)$ of the diagram is exactly $r(\sigma)$ for some negative $(q + 1)$ -simplex σ . Therefore, $n\mu_n(U_M)$ is equal to $N_q(\mathbb{X}_n, M)$, the number of negative $(q + 1)$ -simplexes in the filtration $\mathcal{K}(\mathbb{X}_n)$ appearing after $M_n := n^{-1/d}M$. More details about this pairing between simplexes of the filtration can, for instance, be found in [20], section VIII.1.

(ii) The number of negative simplexes in the Čech and Rips filtration can be efficiently bounded thanks to elementary geometric arguments.

6.4.1. Geometric arguments for the Rips filtration. We have

$$(6.7) \quad N_q(\mathbb{X}_n, M) = \frac{1}{q+2} \sum_{i=1}^n \#\Xi(x_i, \mathbb{X}_n),$$

where, for $x \in \mathbb{X}$, with \mathbb{X} a finite set, $\Xi(x, \mathbb{X})$ is the set of negative $(q + 1)$ -simplexes (and therefore of size $q + 2$) in $\mathcal{R}(\mathbb{X})$ that are containing x , and have a filtration time larger than M_n . The following construction is inspired by the proof of Lemma 2.4 in [27].

The angle (with respect to 0) of two vectors $x, y \in \mathbb{R}^d$ is defined as

$$\angle xy := \arccos \left(\frac{\langle x, y \rangle}{\|x\| \|y\|} \right).$$

The angular section of a cone A is defined as $\sup_{x, y \in A} \angle xy$. Denote by $C(x, r)$ the cube centered at x of side length $2r$. For $0 < \delta < 1$, and for each face of the cube $C(x, r)$, consider a regular grid with spacing δr , so that the center of each face is one of the grid points. This results in a partition of the boundary of the cube $C(x, r)$ into $(d - 1)$ -dimensional cubes $(C_j^\delta(x, r))_{j=1 \dots Q}$ of side length δr . Using this partition of the boundary of $C(x, r)$, we construct a partition of $C(x, r)$ into closed convex cones $(A_j^\delta(x, r))_{j=1 \dots Q}$, where each cone $A_j^\delta(x, r)$ is defined as a d -simplex spanned by x and one of the $(d - 1)$ -dimensional cubes $C_j^\delta(x, r)$ of side length δr on a face of $C(x, r)$. In other words, the point x is the

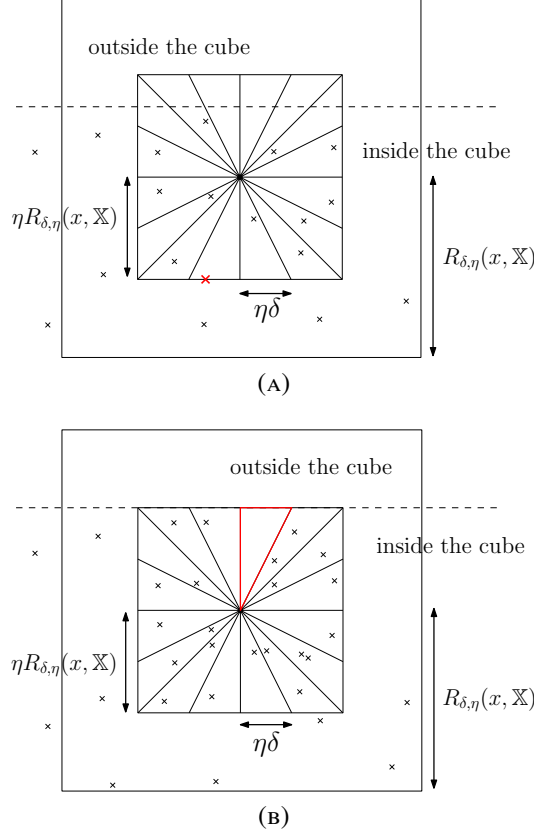


FIGURE 3. Illustration of the definition of $R := R_{\delta, \eta}(x, \mathbb{X})$ for some two point clouds \mathbb{X} . The dashed line indicates the boundary of $[0, 1]^d$. In Figure 3a, the radius R is such that there is a point (indicated in red) on $C_j^\delta(x, R)$. In Figure 3b, there is a cone $A_j^\delta(x, R)$, indicated in red, for which $C_j^\delta(x, R)$ is on some face of the cube $[0, 1]^d$.

apex of each $A_j^\delta(x, r)$, and $C_j^\delta(x, r)$ is its base. We call two such cones $A_j^\delta(x, r)$ and $A_{j'}^\delta(x, r)$ adjacent, if $A_j^\delta(x, r) \cap A_{j'}^\delta(x, r) \neq \{x\}$.

Fix $0 < \eta < 1$, and define $R_{\delta, \eta}(x, \mathbb{X}_n)$ to be the smallest radius r so that each cone $A_j^\delta(x, \eta r)$ in $C(x, \eta r)$ either contains a point of \mathbb{X}_n other than x , or is not a subset of $(0, 1)^d$ (see Figure 3 for an illustration).

Lemma 6.1. *Let $x \in [0, 1]^d$. Fix $\delta > 0$, and $0 < r \leq \frac{1}{2}$, and let $A_j^\delta(x, r)$ be a cone of $C(x, r)$ whose base $C_j^\delta(x, r)$ intersects $[0, 1]^d$. Then, either $A_j^\delta(x, r)$ is a subset of $[0, 1]^d$, or there exists a cone $A_{j'}^\delta(x, r)$ of $C(x, r)$ adjacent to $A_j^\delta(x, r)$ that is a subset of $[0, 1]^d$.*

Proof. A necessary and sufficient condition for a cone $A_j^\delta(x, r)$ to be a subset of $[0, 1]^d$ is that $C_j^\delta(x, r) \subset [0, 1]^d$. Suppose that this is not the case, i.e. we have $C_j^\delta(x, r) \setminus [0, 1]^d \neq \emptyset$. For each coordinate $i = 1, \dots, d$ for which $C_j^\delta(x, r)$

extends beyond a face of $[0, 1]^d$, move one step in the ‘opposite’ direction, and find the corresponding adjacent cone. The fact that $r \leq 1/2$ ensures that these (at most d) steps, each of size $r\delta$, do not make the exterior boundary of the corresponding adjacent cone extend beyond any of the opposite faces of the cube corresponding to the directions of the steps. \square

Note that the angular section (with respect to x) of the union of a cone $A_j^\delta(x, r)$ and its adjacent cones is bounded by $c\delta$ for some constant c .

Lemma 6.2. *Let $\eta = \min\{1/\sqrt{d}, 1/2\}$. There exists a $\delta = \delta(d) > 0$, such that each simplex σ of $\Xi(x, \mathbb{X}_n)$ is included in $C(x, R_{\delta, \eta}(x, \mathbb{X}_n))$. Furthermore, $\Xi(x, \mathbb{X}_n)$ is empty if $R_{\delta, \eta}(x, \mathbb{X}_n) > M_n$.*

Proof. To ease notation, denote $R_{\delta, \eta}(x, \mathbb{X}_n)$ by R . We are going to prove that all negative simplexes containing x are included in $C(x, R)$, a fact that proves the two assertions of the lemma. First, if $\eta R \geq 1/2$, then $C(x, R)$ contains $[0, 1]^d$ and the result is trivial. So, assume that $\eta R < 1/2$, and consider a $(q+1)$ -simplex $\sigma = \{x, x_1, \dots, x_{q+1}\}$ that is not contained in $C(x, R)$. Assume without loss of generality that x_1 is the point in σ maximizing the distance to x , which in particular means that x_1 is not in $C(x, R)$. The line $[x, x_1]$ hits $C(x, \eta R)$ at some cone $A_j^\delta(x, \eta R)$. By Lemma 6.1 and the definition of R , if A^δ denotes the union of $A_j^\delta(x, \eta R)$ and its adjacent cones in $C(x, \eta R)$, then there exists a point z of \mathbb{X}_n in $A^\delta \subset C(x, \eta R)$ and the angle $\angle xzx_1$ formed by $[z, x]$ and $[z, x_1]$ is smaller than $c\delta$. Let us prove that all the $(q+1)$ -simplexes σ_t of the form $(\sigma \setminus \{t\}) \cup \{z\}$, for $t \in \sigma$, have a filtration time smaller than $r(\sigma)$. If this is the case, then the cycle formed by the σ_t ’s and σ is contained in the complex at time $r(\sigma)$, meaning that σ is not negative, concluding the proof. Therefore, it suffices to prove that $\|z - x\| \leq r(\sigma)$ and that $\|z - x_i\| \leq r(\sigma)$ for all i :

- $\|z - x\| \leq \sqrt{d}\eta R \leq \|x - x_1\|$.
- If $\angle xzx_1 < c\delta \leq \pi/3$,

$$\begin{aligned} \|z - x_1\|^2 &= \|z - x\|^2 + \|x_1 - x\|^2 - 2\langle z - x, x_1 - x \rangle \\ &\leq \|z - x\|^2 + \|x_1 - x\|^2 - \|z - x\|\|x_1 - x\| \leq \|x_1 - x\|^2 \leq r(\sigma)^2. \end{aligned}$$

For $i \geq 2$, we have $\|x - x_i\| \leq \|x - x_1\|$ by assumption. Let $I(z)$ denote the set of all $t \in \mathbb{R}^d$ with $\|z - t\| \geq \|x - t\|$ and $\|z - t\| \geq \|x_1 - t\|$, i.e. $I(z)$ is the intersection of two half spaces (see Figure 4). Let $F_x(z) = d(I(z), x)$. If we find a δ with $F_x(z) > \|x - x_1\|$ for all $z \in A^\delta$, then no x_i is in $I(z)$, whatever the position of $z \in A^\delta$ is, meaning that all x_i ’s satisfy $\|z - x_i\| \leq \max\{\|x - x_i\|, \|x_1 - x_i\|\} \leq r(\sigma)$, concluding the proof. The method of Lagrange multipliers shows that $F_x(z)^2$ is a continuous function of z , with a known (but complicated) expression. A straightforward study of this expression shows that for δ small enough, the minimum of F_x on A^δ can be made arbitrarily large: therefore, there exists δ such that $F_x(z) > \sqrt{d} \geq \|x - x_1\|$, for all $z \in A^\delta$. \square

6.4.2. Construction for the Čech filtration. A similar construction works for the Čech filtration, but the arguments are slightly different. First, note that each

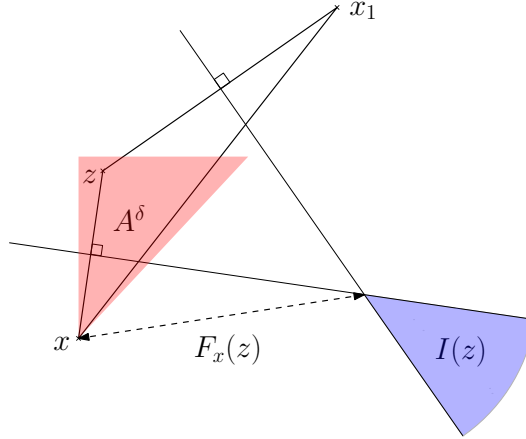


FIGURE 4. The geometric construction used in the proof of Lemma 6.2. The red region represents A^δ whereas the blue region represents $I(z)$ for some point z in A^δ . If δ is made sufficiently small, the distance $F_x(z)$ between x and $I(z)$ can be made arbitrarily large.

negative simplex σ in the Čech filtration is such that there exists a subsimplex σ' of σ that enters in the filtration at the same time $r(\sigma)$ as σ , and so that $r(\sigma)$ is the circumradius of σ' . Then

$$\begin{aligned} N_q(\mathbb{X}_n, M) &= \sum_{\sigma \in C_{q+1}(\mathbb{X}_n)} \mathbf{1}\{\sigma \text{ negative}, r(\sigma) \geq M\} \\ &= \sum_{\sigma \in C_{q+1}(\mathbb{X}_n)} \frac{1}{\#\sigma'} \sum_{i=1}^n \mathbf{1}\{\sigma \text{ negative}, r(\sigma) \geq M, X_i \in \sigma'\} \\ &\leq \sum_{i=1}^n \#\Xi'(X_i, \mathbb{X}_n), \end{aligned}$$

where $\Xi'(x, \mathbb{X}_n)$ the set of negative $(q+1)$ -simplexes σ in the Čech filtration $C(\mathbb{X}_n)$ with $r(\sigma) \geq M$ and $x \in \sigma'$.

Lemma 6.3. For $\eta = \min\{1/\sqrt{d}, 1/2\}$ and some $\delta = \delta(d) > 0$, each simplex σ of $\Xi'(x, \mathbb{X}_n)$ is included in $C(x, R_{\delta, \eta}(x, \mathbb{X}_n))$. Furthermore, $\Xi'(x, \mathbb{X}_n)$ is empty if $R_{\delta, \eta}(x, \mathbb{X}_n) > M_n$.

Proof. Recall the definition of $C(x, r)$ and the partition of $C(x, r)$ into the cones $(A_j^\delta(x, r))_{j=1 \dots Q}$ with corresponding bases $(C_j^\delta(x, r))_{j=1 \dots Q}$. As above, denote $R_{\delta, \eta}(x, \mathbb{X}_n)$ by R . Let $\sigma = \{x, x_1, \dots, x_{q+1}\}$ denote a $(q+1)$ -simplex not included in $C(x, R)$, with $r(\sigma) \geq M$. As in the Rips case, the result is trivial if $\eta R \geq 1/2$. By definition of the Čech filtration, the intersection $\bigcap_{i=0}^{q+1} \overline{B}(x_i, r(\sigma))$ consists of a singleton $\{y\}$. If there is a point z of \mathbb{X}_n in $B(y, r(\sigma))$, then, by the nerve theorem applied to $\sigma \cup \{z\}$, we can conclude with similar arguments as in Lemma 6.2 that σ is positive in the filtration, meaning that every negative $\sigma \in \Xi(x, \mathbb{X}_n)$ has to be included in $C(x, R)$.

Let us prove the existence of such a z . As $x \in \sigma'$, the distance between x and y is equal to $r(\sigma) \geq R$. Therefore, the line $[x, y]$ hits $C(x, \eta R)$ in some cone $A_j^\delta(x, \eta R)$, whose base $C_j^\delta(x, r)$ intersects $[0, 1]^d$, as it intersects $[x_0, y]$. As in the Rips case, there exists a point z of \mathbb{X}_n in $C(x, \eta R)$ such that the angle made by z , x and y is smaller than $c\delta$. As before, it can then be argued that $\|y - z\| \leq r(\sigma)$, concluding the proof. \square

Remark 6.4. Note that the fact that the support of κ is the cube only enters the picture through the geometric arguments used here and in the above proof. Some more refine work is needed to show that a similar construction holds when the cube is replaced by a convex body.

. In the following, fix $\eta = \min\{1/\sqrt{d}, 1/2\}$, choose δ sufficiently small, and let $R_{\delta, \eta}(x, \mathbb{X}_n)$, $A_j^\delta(x, r)$ and $C_j^\delta(x, r)$ be denoted by $R(x, \mathbb{X}_n)$, $A_j(x, r)$ and $C_j(x, r)$ respectively. Both $\Xi(x, \mathbb{X}_n)$ and $\Xi'(x, \mathbb{X}_n)$ are included in the set of $(q+1)$ -tuples of $\mathbb{X}_n \cap C(x, R(x, \mathbb{X}_n))$, so that the following inequality holds for either the Rips or the Čech filtration:

$$(6.8) \quad N_q(\mathbb{X}_n, M) \leq \sum_{i=1}^n \mathbf{1}\{R(X_i, \mathbb{X}_n) > M_n\} (\#\mathbb{X}_n \cap C(X_i, R(X_i, \mathbb{X}_n)))^{q+1}.$$

Denote $R(X_1, \mathbb{X}_n)$ by R_n . As we will see, an estimate of the tail of R_n is sufficient to get a control of $N_q(\mathbb{X}_n, M)$. The probability $P(R_n > t)$ is bounded by the probability that one of the cones pointing at X_1 , of radius $t/2$, wholly included in the cube $[0, 1]^d$, is empty. Conditionally on X_1 , this probability is exactly the probability that a binomial process with parameters $n-1$ and κ does not intersect this cone. Therefore,

$$(6.9) \quad P(R_n > t) \leq c \exp(-cnt^d),$$

and we obtain, for $\lambda > 0$,

$$(6.10) \quad \begin{aligned} E \left[e^{\lambda n R_n^d} \right] &= \int_1^\infty P(\lambda n R_n^d > \ln(t)) dt \\ &\leq \int_1^\infty c \exp\left(-c \frac{\ln(t)}{\lambda}\right) dt < \infty \text{ if } \lambda < c/2. \end{aligned}$$

Lemma 6.5. *The random variable $\#\mathbb{X}_n \cap C(X_1, R_n)$ has exponential tail bounds: for $t > 0$,*

$$P(\#\mathbb{X}_n \cap C(X_1, R_n) > t) \leq c \exp(-ct).$$

Proof. Conditionally on X_1 and R_n , two possibilities may occur. In the first one, the cube centered at X_1 of radius ηR_n contains a point on its boundary, in the cone $A_{j_0}(X_1, \eta R_n)$. Denote this event E and let Q_0 be the number of cones wholly included in the support. The configuration of \mathbb{X}_n is a binomial process conditioned to have at least one point in the cones $A_j(X_1, \eta R_n)$ wholly included in the cube, except for $j = j_0$, and a point on $C_{j_0}(X_1, \eta R_n)$. In this

case, $\#(\mathbb{X}_n \cap C(X_1, R_n))$ is equal to $Q_0 + Z$, where Z is a binomial variable of parameters $n - Q_0$ and

$$\int_{C(X_1, R_n) \setminus A_{j_0}(X_1, \eta R_n)} \kappa(x) dx \leq cR_n^d.$$

Therefore, for $\beta > 0$, using a Chernoff bound and a classical bound on the moment generating function of a binomial variable:

$$\begin{aligned} P(\#(\mathbb{X}_n \cap C(X_1, R_n)) > t | R_n, Q_0, E) &\leq P(Q_0 + Z > t | R_n, Q_0) \\ &\leq \frac{e^{\beta Q_0} E[e^{\beta Z} | R_n]}{e^{\beta t}} \\ &\leq \frac{e^{\beta Q_0} e^{cnR_n^d(e^\beta - 1)}}{e^{\beta t}}, \end{aligned}$$

where Q is the number of elements in the partition of $C(x, R)$. Take β sufficiently small so that $E[e^{cnR_n^d(e^\beta - 1)}] < \infty$ (such a β exists by equation (6.10)). We have the conclusion in this first case.

The other possibility is that there exists a cone not wholly included in the cube containing no point of \mathbb{X}_n . In this case, the configuration of \mathbb{X}_n is a binomial process conditioned on having at least one point in the cones $A_j(X_1, R_n)$ wholly included in cube and no point in a certain cone not wholly included in the cube. Likewise, a similar bound is shown. \square

We are now able to finish the proof of Proposition 4.6: for $p \geq 1$,

$$\begin{aligned} E[\mu_n(U_M)^p] &= E \left[\left(\frac{N_q(\mathbb{X}_n, M)}{n} \right)^p \right] \\ &\leq \frac{1}{(q+2)^p} E \left[\left(\frac{\sum_{i=1}^n \mathbf{1}\{R(X_i, \mathbb{X}_n) > M_n\} (\#(\mathbb{X}_n \cap C(X_i, R(X_i, \mathbb{X}_n))))^{q+1}}{n} \right)^p \right] \\ &\leq \frac{1}{(q+2)^p} E \left[\mathbf{1}\{R_n > M_n\} \#(\mathbb{X}_n \cap C(X_1, R_n))^{p(q+1)} \right] \text{ by Jensen's inequality} \\ &\leq \frac{1}{(q+2)^p} \left(P(R_n > M_n) E[\#(\mathbb{X}_n \cap C(X_1, R_n))^{2p(q+1)}] \right)^{1/2}. \end{aligned}$$

Lemma 6.5 implies that, for $p' > 0$,

$$\begin{aligned} E \left[\#(\mathbb{X}_n \cap C(X_1, R_n))^{p'} \right] &\leq \int_1^\infty c e^{-ct^{1/p'}} dt \\ &= \frac{p'c}{C^{p'}} \int_1^\infty u^{p'-1} e^{-u} du = \frac{c}{C^{p'}} (p')!. \end{aligned}$$

Therefore, for $q \geq 1$,

$$\begin{aligned} E[\mu_n(U_M)^{p/(q+1)}] &\leq c \frac{1}{(q+2)^{p/(q+1)}} \exp(-cM^d) \left((2p)! C^{-2p} \right)^{1/2} \\ &\leq \exp(-cM^d) c^p p!. \end{aligned}$$

To finish the proof, we use a simple lemma relating the moments of a random variable to its tail.

Lemma 6.6. *Let X be a positive random variable such that there exists constants $A, C > 0$ with*

$$(6.11) \quad E[X^k] \leq AC^k k!.$$

Then, there exists a constant $c > 0$ such that $\forall x > 0, P(X > x) \leq A \exp(-cx)$.

Proof. Fix $\lambda = \frac{1}{2C}$. The moment generating function of X in λ is bounded by:

$$E[e^{\lambda X}] = \sum_{k \geq 0} \frac{\lambda^k E[X^k]}{k!} \leq A \sum_{k \geq 0} \lambda^k C^k = A.$$

Therefore, using a Chernoff bound, $P(X > x) \leq A \exp(-\lambda x)$. \square

Apply Lemma 6.6 to $X = \mu_n(U_M)^{1/(q+1)}$ to obtain the assertion of Proposition 4.6. \square

6.5. Proof of Theorem 4.4.

6.5.1. *Step 1: Convergence for functions vanishing on the diagonal.* The first step of the proof is to show that the convergence holds $C_0(\mathbb{R}_>^2)$, the set of continuous bounded functions vanishing on the diagonal. The crucial part of the proof consists in using Proposition 4.6, which bounds the total number of points in the diagrams. An elementary lemma from measure theory is then used to show that it implies the a.s. convergence for vanishing functions. We say that a sequence of measures $(\mu_n)_{n \geq 0}$ converges C_0 -vaguely to μ if $\mu_n(\phi) \rightarrow \mu(\phi)$ for all functions ϕ in $C_0(\mathbb{R}_>^2)$.

Lemma 6.7. *Let E be a locally compact Hausdorff space. Let $(\mu_n)_{n \geq 0}$ be a sequence of Radon measure on E which converges C_c -vaguely to some measure μ . If $\sup_n |\mu_n| < \infty$, then $(\mu_n)_{n \geq 0}$ converges C_0 -vaguely to μ .*

Proof. Let (h_q) be a sequence of functions with compact support converging to 1 and let $\phi \in C_0(E)$. Fix $\varepsilon > 0$. By definition of $C_0(E)$, there exists a compact set K_ε such that ϕ is smaller than ε outside of K_ε . For q large enough, the support of h_q includes K_ε . Let $\phi_q = \phi \cdot h_q$. Then,

$$\begin{aligned} |\mu_n(\phi) - \mu(\phi)| &\leq |\mu_n(\phi) - \mu_n(\phi_q)| + |\mu_n(\phi_q) - \mu(\phi_q)| + |\mu(\phi_q) - \mu(\phi)| \\ &\leq (\sup_n |\mu_n| + |\mu|)\varepsilon + |\mu_n(\phi_q) - \mu(\phi_q)|. \end{aligned}$$

As $(\mu_n)_n$ converges vaguely to μ , the last term of the sum converges to 0 when ε is fixed. Hence,

$$\limsup_{n \rightarrow \infty} |\mu_n(\phi) - \mu(\phi)| \leq \left(\sup_n |\mu_n| + |\mu| \right) \varepsilon.$$

As this holds for all $\varepsilon > 0$, $\mu_n(\phi)$ converges to $\mu(\phi)$. \square

Taking $M = 0$ in Proposition 4.6, we see that $\sup_n E[|\mu_n|] < \infty$. Therefore, the C_0 -vague convergence of $E[\mu_n]$ is shown in the binomial setting. To show that the convergence also holds almost surely for $|\mu_n|$, we need to show that $\sup_n |\mu_n| < \infty$. For this, we use concentration inequalities. We do not show concentration inequalities for $|\mu_n|$ directly. Instead, we derive concentration inequalities for $\sum_{i=1}^n \#(\mathbb{X}_n \cap C(X_i, R(X_i, \mathbb{X}_n)))^{q+1}$, which is a majorant of $|\mu_n|$. Recall that $R(X_i, \mathbb{X}_n)$ is defined as the smallest radius R such that, for some fixed parameter $\eta > 0$, and for each $j = 1 \dots Q$, $A_j(X_i, \eta R)$, either contains a point of \mathbb{X}_n different than X_i , or is not contained in the cube. To ease the notations, we denote $R(X_i, \mathbb{X}_n)$ by $R_{i,n}$.

Lemma 6.8. *Fix $M \geq 0$ and define $Z_n^M = \sum_{i=1}^n \#(\mathbb{X}_n \cap C(X_i, R_{i,n}))^q \mathbf{1}\{R_{i,n} \geq M\}$. Then, for every $\varepsilon > 0$, there exists a constant $c_\varepsilon > 0$ such that*

$$(6.12) \quad P(|Z_n^M - E[Z_n^M]| > t) \leq \frac{n^{\frac{3}{2}+\varepsilon}}{t^3} \exp(-c_\varepsilon n^{-1} M^d).$$

The constant c_ε depends on ε, d, q and κ .

As a consequence of the concentration inequality, $n^{-1}Z_n^0$ is almost surely bounded. Indeed, choose $\varepsilon < 1/2$:

$$P(n^{-1}|Z_n^0 - E[Z_n^0]| > t) \leq \frac{n^{\frac{3}{2}+\varepsilon}}{(nt)^3} = \frac{n^{-\frac{3}{2}+\varepsilon}}{t^3}.$$

By Borel-Cantelli lemma, almost surely, for n large enough, $n^{-1}|Z_n^0 - E[Z_n^0]| \leq 1$. Moreover, $\sup_n n^{-1}E[Z_n^0]$ is finite. As a consequence, $\sup_n n^{-1}Z_n^0$ is almost surely finite. As this is an upper bound of $\sup_n |\mu_n|$, we have proven that $\sup_n |\mu_n| < \infty$ almost surely. By Lemma 6.7, the sequence μ_n converges C_0 -vaguely to μ . The proof of Lemma 6.8 is based on an inequality of the Efron-Stein type and is rather long and technical. It can be found in Section 6.7.

We now briefly consider the Poisson setting. Define $\mu'_n = \mu_{N_n} \times (N_n/n)$, where $(N_i)_{i \geq 1}$ is some sequence of independent Poisson variables of parameter n , independent of $(X_i)_{i \geq 1}$.

- $E[|\mu'_n|] = E\left[\frac{N_n}{n} E[|\mu_{N_n}| | N_n]\right] \leq \sup_n E[|\mu_n|] < \infty$. Therefore, C_0 convergence of the expected diagram holds in the Poisson setting.

- Likewise, it is sufficient to show that $\sup_n \frac{N_n}{n} < \infty$ to conclude to the C_0 convergence of the diagram in the Poisson setting. Fix $t > 1$. It is shown in the chapter 1 of the monograph [29] that $P(N_n > nt) \leq \exp(-nH(t))$, where $H(t) = 1 - t + t \ln t$. This gives us

$$(6.13) \quad \begin{aligned} P\left(\sup_n \frac{N_n}{n} \leq t\right) &= \prod_{n \geq 0} (1 - P(N_n > nt)) \\ &\geq \prod_{n \geq 0} (1 - \exp(-nH(t))) = \exp\left(\sum_{n \geq 0} \ln(1 - \exp(-nH(t)))\right) \end{aligned}$$

The series $\sum_n \ln(1 - x^n)$ is equal to $-\sum_n \sigma(n)/nx^n$ when $|x| < 1$, and where $\sigma(n)$ is the sum of the proper divisors of n . Therefore it is a power series, and

is continuous on $] - 1, 1[$. Since t tends to infinity, $\exp(-H(t))$ converges to 0, and thus the quantity appearing in the right hand side of (6.13) converges to 1 as t tends to infinity.

6.5.2. Step 2: Convergence for functions with polynomial growth. The second step consists in extending the convergence to functions $\phi \in C_{\text{poly}}(\mathbb{R}_{>}^2)$. We only show the result for binomial processes. The proof can be adapted to the Poisson case using similar techniques as at the end of Step 1. The core of the proof is a bound on the number of points in a diagram with high persistence. For $M > 0$, define $T_M = \{r = (r_1, r_2) \in \mathbb{R}_{>}^2 \text{ s.t. pers}(r) \geq M\}$. Let $P_n(M) = n\mu_n(T_M)$ denote the number of points in the diagram with persistence larger than M .

First, we show that the expectation of $P_n(M)$ converges to 0 at an exponential rate when M tends to ∞ . The random variable $P_n(M)$ is bounded by $n\mu_n(U_M)$. By Proposition 4.6, recalling that q is the degree of homology,

$$\begin{aligned}
 E[P_n(M)] &\leq \int_0^\infty P(n\mu_n(U_M) \geq t) dt \\
 &\leq \int_0^\infty c \exp(-c(M^d + (t/n)^{1/(q+1)})) dt \\
 (6.14) \quad &\leq cn \exp(-cM^d) \int_0^\infty \exp(-u) qu^q du = cn \exp(-cM^d)
 \end{aligned}$$

Fix a sequence (g_M) of continuous functions with support inside the complement of T_M taking their values in $[0, 1]$, equal to 1 on T_{M-1}^c . Let ϕ be a function with polynomial growth, i.e. satisfying (4.6) for some $A, \alpha > 0$. Define $\phi_M = \phi \cdot g_M$. We have the decomposition:

$$(6.15) \quad E[\mu_n(\phi)] = (E[\mu_n(\phi)] - E[\mu_n(\phi_M)]) + E[\mu_n(\phi_M)].$$

As $\phi_M \in C_0(\mathbb{R}_{>}^2)$, the second term on the right converges to $\mu(\phi_M)$. The first term on the right is bounded by

$$\begin{aligned}
 E[\mu_n(\phi)] - E[\mu_n(\phi_M)] &\leq E[\mu_n(A(1 + \text{pers}^\alpha)(1 - g_M))] \\
 &\leq AE[P_n(M)]/n + AE[\text{Pers}_\alpha(M; \text{dgm}[\mathbb{X}_n])]/n \\
 (6.16) \quad &\leq cA \exp(-cM^d) + AE[\text{Pers}_\alpha(M; \text{dgm}[\mathbb{X}_n])]/n,
 \end{aligned}$$

using inequality (6.14). It is shown in [16] that

$$\text{Pers}_\alpha(M; \text{dgm}[\mathbb{X}_n]) \leq M^\alpha P_n(M) + \alpha \int_M^\infty P_n(\varepsilon) \varepsilon^{\alpha-1} d\varepsilon.$$

Hence, by Fubini's theorem and inequality (6.14):

$$(6.17) \quad E[\text{Pers}_\alpha(M; \text{dgm}[\mathbb{X}_n])]/n \leq cM^\alpha \exp(-cM^d) + c\alpha \int_M^\infty \exp(-c\varepsilon^d) \varepsilon^{\alpha-1} d\varepsilon.$$

and this quantity goes to 0 as M goes to infinity. Moreover, applying this inequality to $M = 0$, we get that $C_0 = \sup_n E[\mu_n(\phi)] < \infty$. Therefore, $\lim_{n \rightarrow \infty} E[\mu_n(\phi_M)] = \mu(\phi_M) \leq C_0$. By the monotone convergence theorem, $\mu(\phi_M)$ converges to $\mu(\phi)$ when ϕ is non negative, with $\mu(\phi)$ finite by the latter

inequality. If ϕ is not always non negative, we conclude by separating its positive and negative parts. Finally, looking at the bounds (6.16) and (6.17),

$$\limsup_{n \rightarrow \infty} |E[\mu_n(\phi)] - \mu(\phi)| \leq \limsup_{n \rightarrow \infty} (E[\mu_n(\phi)] - E[\mu_n(\phi_M)]) + |\mu(\phi_M) - \mu(\phi)|,$$

which converges to 0 as $M \rightarrow \infty$.

We now prove that $\mu_n(\phi) - E[\mu_n(\phi)]$ converges a.s. to 0. Similar to the above, it is enough to show that $P_n(M)$ is almost surely bounded by a quantity independent of n , which converges to 0 at an exponential rate when M goes to ∞ . The random variable $(q+2)P_n(M)$ is bounded by Z_n^M , which is defined in Lemma 6.8, and whose expectation is controlled. Therefore, it remains to show that Z_n^M is close to its expectation. We have

$$\begin{aligned} \limsup_{n \rightarrow \infty} |\mu_n(\phi) - E[\mu_n(\phi)]| &\leq \limsup_{n \rightarrow \infty} (|\mu_n(\phi - \phi_M)| + |(\mu_n - E[\mu_n])(\phi_M)| \\ &\quad + |E[\mu_n](\phi - \phi_M)|) \\ (6.18) \quad &\leq \limsup_{n \rightarrow \infty} |\mu_n(\phi - \phi_M)| + 0 + cA \exp(-cM^d) \\ &\quad + cM^\alpha \exp(-cM^d) + c\alpha \int_M^\infty \exp(-c\varepsilon^d) \varepsilon^{\alpha-1} d\varepsilon \end{aligned}$$

by inequalities (6.16) and (6.17). The random variable $|\mu_n(\phi - \phi_M)|$ is bounded by

$$\begin{aligned} &An^{-1} \left(P_n(M) + M^\alpha P_n(M) + \alpha \int_M^\infty P_n(\varepsilon) \varepsilon^{\alpha-1} d\varepsilon \right) \\ (6.19) \quad &\leq An^{-1} \left(Z_n^{M_n} + M^\alpha Z_n^{M_n} + \alpha \int_M^\infty Z_n^{\varepsilon_n} \varepsilon^{\alpha-1} d\varepsilon \right) \end{aligned}$$

where $M_n = n^{-1/d}M$ and $\varepsilon_n = n^{-1/d}\varepsilon$. As a consequence of Lemma 6.8, by choosing ε so that $-3/2 + \varepsilon < -1$,

$$P \left(\sup_n n^{-1} |Z_n^{M_n} - E[Z_n^{M_n}]| > t \right) \leq c \frac{\exp(-cM^d)}{t^3}.$$

Fixing $t = \exp(-(c/6)M^d)$ and using Borel-Cantelli lemma, for $M \in \mathbb{N}$ large enough, $\sup_n n^{-1} |Z_n^{M_n} - E[Z_n^{M_n}]| < \exp(-(c/6)M^d)$. Also, $E[Z_n^{M_n}] \leq nc \exp(-cM^d)$. Therefore, for $\alpha \geq 0$,

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} n^{-1} M^\alpha Z_n^{M_n} = 0 \text{ a.s.}$$

The third term in the sum (6.19) is less straightforward to treat. As Z_n^M is a decreasing function of M , for $M \in \mathbb{N}$ large enough and with $k_n = n^{-1/d}k$:

$$\begin{aligned} \int_M^\infty n^{-1} Z_n^{\varepsilon_n} \varepsilon^{\alpha-1} d\varepsilon &\leq \sum_{k \geq M} n^{-1} Z_n^{k_n} \int_k^{k+1} \varepsilon^{\alpha-1} d\varepsilon \\ &\leq \sum_{k \geq M} c \exp(-ck^d) \frac{1}{\alpha} ((k+1)^\alpha - k^\alpha) \end{aligned}$$

$$\leq \sum_{k \geq M} c \exp(-ck^d) \frac{k^\alpha 2^\alpha}{\alpha} = o(1).$$

As a consequence, $\lim_{M \rightarrow \infty} \limsup_n |\mu_n(\phi - \phi_M)| = 0$. As the three last terms appearing in inequality (6.18) also converges to 0 when M goes to infinity, we have proven that $\mu_n(\phi) - E[\mu_n(\phi)]$ converges a.s. to 0. Therefore, $\mu_n(\phi)$ converges a.s. to $\mu(\phi)$.

Finally, we have to prove assertion (ii) in Theorem 4.4, i.e. that the convergence holds in L_p . As the convergence holds in probability, it is sufficient to show that $(\mu_n(\phi)^p)_n$ is uniformly integrable. Observing that $E[\mu_n(\phi)^p \mathbf{1}\{\mu_n(\phi)^p > M\}] \leq (E[\mu_n(\phi)^{2p}] P(\mu_n(\phi)^p > M))^{1/2}$, uniform integrability follows from $\sup_n E[\mu_n(\phi)^p] < \infty$ for any $p > 1$. We have

$$\mu_n(\phi)^p \leq \mu_n(A(1 + \text{pers}^\alpha))^p \leq 2^{p-1} (A^p |\mu_n|^p + \mu_n(\text{pers}^\alpha)^p),$$

and from Proposition 4.6 we easily obtain that $E[|\mu_n|^p]$ is uniformly bounded. We treat the other part by assuming without loss of generality that p is an integer:

$$\begin{aligned} E[\mu_n(\text{pers}^\alpha)^p] &\leq \frac{1}{n^p} E \left[\left(\int_0^\infty P_n(\varepsilon) \varepsilon^{\alpha-1} d\varepsilon \right)^p \right] \\ &= \frac{1}{n^p} E \left[\int_{[0, \infty]^p} P_n(\varepsilon_1) \cdots P_n(\varepsilon_p) (\varepsilon_1 \cdots \varepsilon_p)^{\alpha-1} d\varepsilon_1 \cdots d\varepsilon_p \right] \\ &= \frac{1}{n^p} \int_{[0, \infty]^p} E[P_n(\varepsilon_1) \cdots P_n(\varepsilon_p)] (\varepsilon_1 \cdots \varepsilon_p)^{\alpha-1} d\varepsilon_1 \cdots d\varepsilon_p \\ &\leq \frac{1}{n^p} \int_{[0, \infty]^p} (E[P_n(\varepsilon_1)^p] \cdots E[P_n(\varepsilon_p)^p])^{1/p} (\varepsilon_1 \cdots \varepsilon_p)^{\alpha-1} d\varepsilon_1 \cdots d\varepsilon_p \\ &\leq \frac{1}{n^p} \int_{[0, \infty]^p} (n^{p^2} c \exp(-c(\varepsilon_1^d + \cdots + \varepsilon_p^d)))^{1/p} (\varepsilon_1 \cdots \varepsilon_p)^{\alpha-1} d\varepsilon_1 \cdots d\varepsilon_p \\ &= c \int_{[0, \infty]^p} \exp\left(-\frac{c}{p}(\varepsilon_1^d + \cdots + \varepsilon_p^d)\right) (\varepsilon_1 \cdots \varepsilon_p)^{\alpha-1} d\varepsilon_1 \cdots d\varepsilon_p \\ &= c \left(\int_0^\infty \exp\left(-\frac{c}{p}\varepsilon^d\right) \varepsilon^{\alpha-1} d\varepsilon \right)^p < \infty. \end{aligned}$$

6.6. Proof of Proposition 4.7. The proof relies on the regularity of the number of simplexes appearing at certain scales.

Lemma 6.9. *Let $q \geq 0$. For $r_1 < r_2$, let $F_q(\mathbb{X}_n, r_1, r_2)$ be the number of q -simplexes σ in the filtration $\mathcal{K}(\mathbb{X}_n)$ with $r(\sigma) \in [r_1, r_2]$. Assume that \mathbb{X}_n is a binomial n -sample of density κ . Then,*

$$(6.20) \quad n^{-1} F_q(n^{1/d} \mathbb{X}_n, r_1, r_2) \xrightarrow[n \rightarrow \infty]{L_2} F_q(r_1, r_2),$$

where $F_q(r_1, r_2) \leq cr_2^{2dq-1} |r_2 - r_1|$.

Proof. For a finite set $\mathbb{X} \subset \mathbb{R}^d$, define

$$\xi^{r_1, r_2}(x, \mathbb{X}) = \frac{1}{q+1} \sum_{\sigma \in \mathcal{K}_q(\mathbb{X})} \mathbf{1}\{r(\sigma) \in [r_1, r_2] \text{ and } x \in \sigma\}.$$

Then, $F_q(\mathbb{X}_n, r_1, r_2) = \sum_{x \in \mathbb{X}} \xi^{r_1, r_2}(x, \mathbb{X}_n)$. The paper [30] shows convergence in L_2 of such functionals $\xi(x, \mathbb{X})$ under two conditions. The first one of them is called stabilization. Let \mathbb{P} be a homogeneous Poisson process in \mathbb{R}^d . A quantity $\xi(x, \mathbb{X})$ is stabilizing if, with probability one, there exists some random radius $R < \infty$ such that, for all finite sets A which are equal to \mathbb{P} on $B(0, R)$,

$$\xi(0, \mathbb{P} \cap B(0, R)) = \xi(0, A),$$

Denote this quantity by $\xi_\infty(\mathbb{P})$. In our case, ξ^{r_1, r_2} is stabilizing with $R = 2r_2$. The second condition is a moment condition: there exists some number $\beta > 2$ such that

$$\sup_n E[\xi(n^{1/d} X_1, n^{1/d} \mathbb{X}_n)^\beta] < \infty.$$

Once again, the functional ξ^{r_1, r_2} possesses this property: the random variable $\xi^{r_1, r_2}(n^{1/d} X_1, n^{1/d} \mathbb{X}_n)$ is bounded by the number of q -simplexes of $\mathcal{K}(\mathbb{X}_n)$ containing X_1 and being included in $B(X_1, 2n^{-1/d} r_2)$. This number of q -simplexes is bounded by $\#(\mathbb{X}_n \cap B(X_1, 2n^{-1/d} r_2))^q$, which, in turn, is stochastically dominated by a binomial random variable with parameters n and $cn^{-1} r_2^d$. In particular, its moment of order $3q$ is smaller than a constant independent of n . This means that the moment condition is satisfied. Applying the main theorem of [30], convergence (6.20) is obtained, with $F_q(r_1, r_2) = E[\xi_\infty^{r_1, r_2}(\mathbb{P})]$, where $\xi_\infty^{r_1, r_2}(\mathbb{P}) = \xi^{r_1, r_2}(0, \mathbb{P} \cap B(0, 2r_2))$. The set $\mathbb{P} \cap B(0, 2r_2)$ can be expressed as $\{X_1, \dots, X_N\}$, where $(X_i)_{i \geq 0}$ is a sequence of i.i.d. uniform random variables on $B(0, 2r_2)$, and N is an independent Poisson variable with parameter cr_2^d . Therefore,

$$\begin{aligned} E[\xi_\infty^{r_1, r_2}(\mathbb{P})] &= E \left[\sum_{i_1, \dots, i_q} \mathbf{1}\{r(\{0, X_{i_1}, \dots, X_{i_q}\}) \in [r_1, r_2]\} \right] \\ &= E \left[\frac{N!}{(N-q)!} \right] P(r(\{0, X_1, \dots, X_q\}) \in [r_1, r_2]) \\ &\leq cr_2^{2dq-1} |r_2 - r_1|. \end{aligned}$$

The last inequality is a consequence of (i) the fact that the q -th factorial moment of N equals cr_2^{dq} , and (ii) Theorem 7 in [11]. Indeed, write

$$\begin{aligned} &P(r(\{0, X_1, \dots, X_q\}) \in [r_1, r_2]) \\ &= \int_{B(0, 2r_2)^q} \mathbf{1} \left\{ r \left(\left\{ 0, \frac{x_1}{r_2}, \dots, \frac{x_q}{r_2} \right\} \right) \in \left[\frac{r_1}{r_2}, 1 \right] \right\} dx \\ &= r_2^{dq} \int_{B(0, 2)^q} \mathbf{1} \left\{ r(\{0, u_1, \dots, u_q\}) \in \left[\frac{r_1}{r_2}, 1 \right] \right\} du. \end{aligned}$$

It is shown in [11] that for either the Rips or the Čech filtration built on a random point cloud \mathbb{X} having a smooth density with respect to the Lebesgue measure, the corresponding expected diagrams have a smooth density with respect to the Lebesgue measure on $\mathbb{R}_{>}^2$. With $\mathbb{X} = \{0, X_1, \dots, X_q\}$, a q -sample of uniform random variables on $B(0, 2)$, the diagram $\text{dgm}_q[\mathcal{K}(\mathbb{X})]$ consists of the single point $r = (r_1, r_2)$, with r_2 the filtration time of the simplex \mathbb{X} . Therefore, $r(\{0, X_1, \dots, X_{q-1}\})$ has a smooth density with respect to the Lebesgue measure on \mathbb{R} , which implies the inequality. \square

Fix $0 < r_1 < r_2$. We wish to show that, as r_1 and r_2 get closer, $\pi_1^* \mu([r_1, r_2])$ goes to 0. By the Portemanteau Theorem, $\pi_1^* \mu([r_1, r_2]) \leq \liminf_n \pi_1^* \mu_n([r_1, r_2])$. It is shown in Lemma 6.9 that this quantity is smaller than $cr_2^{2dq-1} |r_2 - r_1|$, a quantity which converges to 0 when r_2 goes to r_1 . A similar proof holds for $\pi_2^* \mu$.

6.7. Proof of Lemma 6.8. The lemma is based on an inequality of the Efron-Stein type, combined with Markov's inequality.

Theorem 6.10 (Theorem 2 in [6]). *Let \mathbb{X} be a measurable set and $F : \mathbb{X}^n \rightarrow \mathbb{R}$ a measurable function. Define a n -sample $\mathbb{X}_n = \{X_1, \dots, X_n\}$ and let $Z = F(\mathbb{X}_n)$. If \mathbb{X}'_n is an independent copy of \mathbb{X}_n , write Z'_i for $F(X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n)$. Define*

$$V = \sum_{i=1}^n E[(Z - Z'_i)^2 | \mathbb{X}_n].$$

Then, for $p \geq 2$, there exists a constant C_p depending only on p such that

$$E[|Z - E[Z]|^p] \leq C_p E[V^{p/2}].$$

Denote $\mathbb{X}_n^i = \mathbb{X}_n \setminus \{X_i\}$ and $S(X_i, \mathbb{X}_n) = \#(\mathbb{X}_n \cap C(X_i, R_{i,n}))^q \mathbf{1}\{R_{i,n} \geq M\}$. We will apply Theorem 6.10 to $F(\mathbb{X}_n) = \sum_{j=1}^n S(X_j, \mathbb{X}_n)$. The quantity $(Z - Z'_i)^2$ is bounded by $2(Z - Z_i)^2 + 2(Z'_i - Z_i)^2$, where $Z_i = F(\mathbb{X}_n^i)$. For most X_j 's, $S(X_j, \mathbb{X}_n) = S(X_j, \mathbb{X}_n^i)$, and therefore V can be efficiently bounded. More precisely,

$$E[V^{p/2}] = E \left[\left(\sum_{i=1}^n E[(Z - Z'_i)^2 | \mathbb{X}_n] \right)^{p/2} \right]$$

(by Jensen's inequality)

$$\begin{aligned} &\leq n^{p/2} E[(Z - Z'_n)^p] \\ &\leq n^{p/2} 2^{p-1} E[(Z - Z_n)^p + (Z'_n - Z_n)^p] \\ &= n^{p/2} 2^p E[(Z - Z_n)^p] \text{ as } (Z, Z_n) \sim (Z'_n, Z_n) \end{aligned}$$

$$= 2^p n^{p/2} E \left[\left(S(X_n, \mathbb{X}_n) + \sum_{j=1}^{n-1} (S(X_j, \mathbb{X}_n) - S(X_j, \mathbb{X}_{n-1})) \right)^p \right]$$

(6.21)

$$\leq 2^p n^{p/2} 2^{p-1} \left(E[S(X_n, \mathbb{X}_n)^p] + E \left[\left(\sum_{j=1}^{n-1} (S(X_j, \mathbb{X}_n) - S(X_j, \mathbb{X}_{n-1})) \right)^p \right] \right).$$

Fix $p = 3$. Lemma 6.5 shows that for $p \geq 1$, $B_p = \sup_n E[S(X_n, \mathbb{X}_n)^p] < \infty$. Define $Y_j = (S(X_j, \mathbb{X}_n) - S(X_j, \mathbb{X}_{n-1}))$. Denote G_j the event that $X_n \in C(X_j, R_{j,n-1})$. If G_j is not realized, then $Y_j = 0$. Expanding the product,

$$(6.22) \quad E \left[\left(\sum_{j=1}^{n-1} (S(X_j, \mathbb{X}_n) - S(X_j, \mathbb{X}_{n-1})) \right)^3 \right] \leq \sum_{j_1, j_2, j_3} E \left[\mathbf{1}_{\{G_{j_1} \cap G_{j_2} \cap G_{j_3}\}} Y_{j_1} Y_{j_2} Y_{j_3} \right] \\ \leq \sum_{j_1, j_2, j_3} P(G_{j_1} \cap G_{j_2} \cap G_{j_3})^{1/q'} E \left[(Y_{j_1} Y_{j_2} Y_{j_3})^{p'} \right]^{1/p'},$$

where $\frac{1}{p'} + \frac{1}{q'} = 1$ and $p' \geq 1$ is some quantity to be fixed later.

- We first bound $E \left[(Y_{j_1} Y_{j_2} Y_{j_3})^{p'} \right]$. If $Y_{j_1} \neq 0$, then $R_{j_1, n-1} > M$. Therefore,

$$E \left[(Y_{j_1} Y_{j_2} Y_{j_3})^{p'} \right] \leq \sqrt{P(R_{j_1, n-1} > M)} E \left[(Y_{j_1} Y_{j_2} Y_{j_3})^{2p'} \right].$$

Also, $E \left[(Y_{j_1} Y_{j_2} Y_{j_3})^{2p'} \right] \leq E \left[Y_{j_1}^{6p'} \right] \leq B_{6p'}$, as $0 \leq Y_{j_1} \leq S(X_{j_1}, \mathbb{X}_n)$. Therefore, using inequality (6.9):

$$E \left[(Y_{j_1} Y_{j_2} Y_{j_3})^{p'} \right] \leq c \exp(-cM^d).$$

- We now bound the probability $P(G_{j_1} \cap G_{j_2} \cap G_{j_3})$.

If $j_1 = j_2 = j_3$, then it is clear that $P(G_{j_1} \cap G_{j_2} \cap G_{j_3}) \leq c/n$. However, in the general case, the joint law of the different $R_{j_i, n-1}$ s becomes of interest. To ease the notation, assume that $j_i = i$ and denote $R_{i, n-1}$ simply by R_i . Also, define D_{ij} the distance between X_i and X_j . The fact that inequality (6.9) still holds conditionally on X_1, X_2 and X_3 , and with the joint laws of R_1, R_2 and R_3 , will be repeatedly used.

Lemma 6.11. *The following bound holds:*

$$(6.23) \quad P(R_1 \geq t_1, R_2 \geq t_2, R_3 \geq t_3 | X_1, X_2, X_3) \leq c \exp(-cn(t_1^d + t_2^d + t_3^d))$$

Proof. Suppose that $\max t_i = t_1$. Inequality (6.9) states that $P(R_1 \geq t_1) \leq c \exp(-cnt_1^d)$. Likewise, it is straightforward to show that a similar bound holds conditionally on X_1, X_2 and X_3 . As $t_1^d \geq \frac{t_1^d + t_2^d + t_3^d}{3}$, the result follows. \square

Let us prove that $P(G_1 \cap G_2) \leq c/n^2$. If the event is realized, then X_n is in the intersection of $C(X_1, R_1)$ and $C(X_2, R_2)$. Therefore, this intersection is non empty and $D_{12} \leq \sqrt{d}(R_1 + R_2)$. Hence,

$$P(G_1 \cap G_2) \leq P(D_{12} \leq \sqrt{d}(R_1 + R_2) \text{ and } X_n \in C(X_1, R_1) \cap C(X_2, R_2)) \\ \leq 2P(D_{12} \leq 2\sqrt{d}R_1 \text{ and } X_n \in C(X_1, R_1)) \\ = 2E \left[\mathbf{1}_{\{D_{12} \leq 2\sqrt{d}R_1\}} P(X_n \in C(X_1, R_1) | \mathbb{X}_{n-1}) \right]$$

$$\begin{aligned}
&\leq 2cE \left[\mathbf{1}\{D_{12} \leq 2\sqrt{d}R_1\} R_1^d \right] \\
&\leq 2cE \left[\int_{(D_{12}/(2\sqrt{d}))^d} P(R_{1,n-1}^d \geq t | X_1, X_2) dt \right] \\
&\leq 2cE \left[\int_{(D_{12}/(2\sqrt{d}))^d} \exp(-cnt) dt \right] \\
&\leq \frac{c}{n} E \left[\exp(-cnD_{12}^d) \right] \\
&= \frac{c}{n} \int_0^1 P(cnD_{12}^d \leq -\ln(t)) dt \\
&\leq \frac{c}{n} \int_0^1 \frac{-\ln(t)}{cn} dt = \frac{c}{n^2}.
\end{aligned}$$

Finally, we bound $P(G_1 \cap G_2 \cap G_3)$. If the event is realized, then

$$\begin{cases} D_{12} \leq \sqrt{d}(R_1 + R_2) \\ D_{23} \leq \sqrt{d}(R_2 + R_3) \\ D_{13} \leq \sqrt{d}(R_1 + R_3) \end{cases} \implies \begin{cases} D_{12} \leq 2\sqrt{d}R_1 \text{ or } D_{12} \leq 2\sqrt{d}R_2 \\ D_{23} \leq 2\sqrt{d}R_2 \text{ or } D_{23} \leq 2\sqrt{d}R_3 \\ D_{13} \leq 2\sqrt{d}R_1 \text{ or } D_{13} \leq 2\sqrt{d}R_3 \end{cases}$$

This last event is an union of eight events. Each of these event is either bounded by an event of the form $(D_{12} \leq 2\sqrt{d}R_1 \text{ and } D_{13} \leq 2\sqrt{d}R_1)$ (six events), or by an event of the form $(D_{12} \leq 2\sqrt{d}R_1 \text{ and } D_{23} \leq 2\sqrt{d}R_2 \text{ and } D_{13} \leq 2\sqrt{d}R_3)$ (two events). Using this, we obtain

$$\begin{aligned}
&P(G_1 \cap G_2 \cap G_3) \\
&\leq 6P(X_n \in C(X_1, R_1) \text{ and } D_{12} \leq 2\sqrt{d}R_1 \text{ and } D_{13} \leq 2\sqrt{d}R_1) \\
&+ 2P(X_n \in C(X_1, R_1) \text{ and } D_{12} \leq 2\sqrt{d}R_1 \text{ and } D_{23} \leq 2\sqrt{d}R_2 \text{ and } D_{13} \leq 2\sqrt{d}R_3) \\
&\leq cE[R_1^d \mathbf{1}\{D_{12} \leq 2\sqrt{d}R_1 \text{ and } D_{13} \leq 2\sqrt{d}R_1\}] \\
&+ cE[R_1^d \mathbf{1}\{D_{12} \leq 2\sqrt{d}R_1 \text{ and } D_{23} \leq 2\sqrt{d}R_2 \text{ and } D_{13} \leq 2\sqrt{d}R_3\}] \\
&= cE \left[\int_{\left(\frac{\max(D_{12}, D_{13})}{2\sqrt{d}}\right)^d}^{\infty} P(R_1^d \geq u | X_1, X_2, X_3) du \right] \\
&+ cE \left[\int_{\left(\frac{D_{12}}{2\sqrt{d}}\right)^d}^{\infty} P \left(R_1^d \geq u \text{ and } R_2 \geq \frac{D_{23}}{2\sqrt{d}} \text{ and } R_3 \geq \frac{D_{13}}{2\sqrt{d}} \middle| X_1, X_2, X_3 \right) du \right] \\
&\leq cE \left[\int_{\left(\frac{\max(D_{12}, D_{13})}{2\sqrt{d}}\right)^d}^{\infty} e^{-cnu} du \right] + cE \left[\int_{\left(\frac{D_{12}}{2\sqrt{d}}\right)^d}^{\infty} e^{(-cn(u+D_{23}^d+D_{13}^d))} du \right] \\
&= \frac{c}{n} E \left[e^{-cn \max(D_{12}, D_{13})^d} \right] + \frac{c}{n} E \left[e^{-cn(D_{12}^d+D_{23}^d+D_{13}^d)} \right]
\end{aligned}$$

$$\begin{aligned} &\leq 2\frac{c}{n}E\left[e^{-cn\max(D_{12},D_{13})^d}\right] = \frac{c}{n}\int_0^1 P\left(\max(D_{12},D_{13})^d \leq \frac{-\ln(t)}{cn}\right) dt \\ &\leq \frac{c}{n}\int_0^1 \left(\frac{-\ln(t)}{cn}\right)^2 dt = \frac{c}{n^3}. \end{aligned}$$

Finally, inequality (6.22) becomes

$$\begin{aligned} E[V^{3/2}] &\leq cn^{3/2}(\exp(-cM^d) + \exp(-cM^d)(n^{1-1/q'} + n^{2-2/q'} + n^{3-3/q'})) \\ &\leq cn^{3/2+3(1-1/q')} \exp(-cM^d). \end{aligned}$$

Choose $p' = 3/\varepsilon$ and apply Markov inequality to conclude.

Acknowledgements. The authors would like to thank the two referees for careful reading of the manuscript. Their very helpful comments lead to a significant improvements of the presentation all over the manuscript.

REFERENCES

- [1] Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., Ziegelmeier, L.: Persistence images: a stable vector representation of persistent homology. *Journal of Machine Learning Research* **18**(8), 1–35 (2017)
- [2] Aronszajn, N.: Theory of reproducing kernels. *Transactions of the American mathematical society* **68**(3), 337–404 (1950)
- [3] Biscio, C.A.N., Møller, J.: The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications. *Journal of Computational and Graphical Statistics* (2019)
- [4] Bobrowski, O., Kahle, M., Skraba, P., et al.: Maximally persistent cycles in random geometric complexes. *The Annals of Applied Probability* **27**(4), 2032–2060 (2017)
- [5] Bobrowski, O., Oliveira, G.: Random čech complexes on riemannian manifolds. *Random Structures & Algorithms* (2017)
- [6] Boucheron, S., Bousquet, O., Lugosi, G., Massart, P., et al.: Moment inequalities for functions of independent random variables. *The Annals of Probability* **33**(2), 514–560 (2005)
- [7] Bubenik, P.: Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research* **16**(1), 77–102 (2015)
- [8] Carriere, M., Cuturi, M., Oudot, S.: Sliced wasserstein kernel for persistence diagrams. In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 664–673. *JMLR. org* (2017)
- [9] Chazal, F., Cohen-Steiner, D., Guibas, L.J., Mémoli, F., Oudot, S.Y.: Gromov-hausdorff stable signatures for shapes using persistence. In: *Computer Graphics Forum*, vol. 28, pp. 1393–1403. *Wiley Online Library* (2009)
- [10] Chazal, F., De Silva, V., Glisse, M., Oudot, S.: *The structure and stability of persistence modules*. Springer (2016)

- [11] Chazal, F., Divol, V.: The Density of Expected Persistence Diagrams and its Kernel Based Estimation. In: 34th International Symposium on Computational Geometry (SoCG 2018), pp. 26:1–26:15 (2018)
- [12] Chazal, F., Fasy, B.T., Lecci, F., Rinaldo, A., Wasserman, L.: Stochastic convergence of persistence landscapes and silhouettes. In: Proceedings of the Thirtieth Annual Symposium on Computational Geometry, p. 474. ACM (2014)
- [13] Chazal, F., Michel, B.: An introduction to topological data analysis: fundamental and practical aspects for data scientists. arXiv:1710.04019 (2017)
- [14] Chen, Y.C., Wang, D., Rinaldo, A., Wasserman, L.: Statistical analysis of persistence intensity functions. arXiv:1510.02502 (2015)
- [15] Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. *Discrete & Computational Geometry* **37**(1), 103–120 (2007)
- [16] Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Mileyko, Y.: Lipschitz functions have L_p -stable persistence. *Foundations of Computational Mathematics* **10**(2), 127–139 (2010)
- [17] Cuevas, A.: Set estimation: Another bridge between statistics and geometry. *Boletín de Estadística e Investigación Operativa* **25**(2), 71–85 (2009)
- [18] Diestel, J.: Sequences and Series in Banach spaces. Graduate Texts in Mathematics. Springer-Verlag (1984)
- [19] Divol, V., Lacombe, T.: Understanding the topology and the geometry of the persistence diagram space via optimal partial transport. arXiv preprint arXiv:1901.03048 (2019)
- [20] Edelsbrunner, H., Harer, J.: Computational topology: an introduction. American Mathematical Soc. (2010)
- [21] Goel, A., Trinh, K.D., Tsunoda, K.: Strong law of large numbers for betti numbers in the thermodynamic regime. *Journal of Statistical Physics* pp. 1–28 (2018)
- [22] Hiraoka, Y., Shirai, T., Trinh, K.D., et al.: Limit theorems for persistence diagrams. *The Annals of Applied Probability* **28**(5), 2740–2780 (2018)
- [23] Kahle, M.: Random geometric complexes. *Discrete & Computational Geometry* **45**(3), 553–573 (2011). DOI 10.1007/s00454-010-9319-3
- [24] Kahle, M., Meckes, E., et al.: Limit theorems for Betti numbers of random simplicial complexes. *Homology, Homotopy and Applications* **15**(1), 343–374 (2013)
- [25] Kusano, G., Fukumizu, K., Hiraoka, Y.: Kernel method for persistence diagrams via kernel embedding and weight factor. *Journal of Machine Learning Research* **18**(189), 1–41 (2018)
- [26] Lee, Y., Barthel, S.D., Dłotko, P., Moosavi, S.M., Hess, K., Smit, B.: Quantifying similarity of pore-geometry in nanoporous materials. *Nature Communications* **8**, 15396 (2017)
- [27] McGivney, K., Yukich, J.E.: Asymptotics for Voronoi tessellations on random samples. *Stochastic Processes and their Applications* **83**(2), 273–288 (1999)

- [28] Nakamura, T., Hiraoka, Y., Hirata, A., Escobar, E.G., Nishiura, Y.: Persistent homology and many-body atomic structure for medium-range order in the glass. *Nanotechnology* **26**(30), 304001 (2015)
- [29] Penrose, M.D.: *Random geometric graphs*. *Oxford Studies of Probability*, 5, Oxford University Press (2003)
- [30] Penrose, M.D., Yukich, J.E.: Weak laws of large numbers in geometric probability. *Annals of Applied Probability* pp. 277–303 (2003)
- [31] Reininghaus, J., Huber, S., Bauer, U., Kwitt, R.: A stable multi-scale kernel for topological machine learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4741–4748 (2015)
- [32] Seversky, L.M., Davis, S., Berger, M.: On time-series topological data analysis: New data and opportunities. In: *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016 IEEE Conference on, pp. 1014–1022. IEEE (2016)
- [33] Skraba, P., Thoppe, G., Yogeshwaran, D.: Randomly weighted d -complexes: Minimal spanning acycles and persistence diagrams. arXiv preprint arXiv:1701.00239 (2017)
- [34] Steele, J.M.: Growth rates of Euclidean minimal spanning trees with power weighted edges. *The Annals of Probability* **16**(4), 1767–1787 (1988)
- [35] Trinh, K.D.: A remark on the convergence of betti numbers in the thermodynamic regime. *Pacific Journal of Mathematics for Industry* **9**(1), 4 (2017)
- [36] Villani, C.: *Optimal transport: old and new*. *Grundlehren der mathematischen Wissenschaften*, 338, Springer Verlag (2008)
- [37] Yao, Y., Sun, J., Huang, X., Bowman, G.R., Singh, G., Lesnick, M., Guibas, L.J., Pande, V.S., Carlsson, G.: Topological methods for exploring low-density states in biomolecular folding pathways. *The Journal of Chemical Physics* **130**(14), 04B614 (2009)
- [38] Yogeshwaran, D., Subag, E., Adler, R.J.: Random geometric complexes in the thermodynamic regime. *Probability Theory and Related Fields* **167**(1), 107–142 (2017). DOI 10.1007/s00440-015-0678-9
- [39] Yukich, J.E.: Asymptotics for weighted minimal spanning trees on random points. *Stochastic Processes and their Applications* **85**(1), 123–138 (2000)
- [40] Yukich, J.E.: *Probability theory of classical Euclidean optimization problems*. Springer (2006)