Topological features of multivariate distributions: dependency on the covariance matrix

Lloyd L. Aromi

University of Barcelona, Barcelona, Spain

Yuri A. Katz

S&P Global Market Intelligence, New York, USA

yuri.katz@spglobal.com

Josep Vives

University of Barcelona, Barcelona, Spain

April 24, 2021

Abstract

Topological data analysis provides a new perspective on many problems in the domain of complex systems. Here, we establish the dependency of the mean value of functional p-norms of 'persistence landscapes' on a uniform scaling of the underlying multivariate distribution. Furthermore, we demonstrate that the average value of p-norms is decreasing, when the covariance in a system is increasing. To illustrate the complex dependency of these topological features on changes of the covariance matrix, we conduct numerical experiments utilizing bi-variate distributions with known statistical properties. Our results help to explain the puzzling behavior of p-norms derived from daily log-returns of major equity indices on European and US markets at the inception phase of the global financial meltdown caused by the COVID-19 pandemic.

Keywords: topological data analysis, statistical topology, complex systems, financial time series

1. Introduction

Topological data analysis (TDA) of noisy data sets combines deterministic constructs of computational topology with statistical and machine learning methods. The latter methods are typically applied by encoding the persistence diagram - one of the main instruments of dimensionality reduction in TDA - into a vector space. The popular way to implement this task is to transform the persistence diagram into the persistent landscape; see Bubenik (2015). This approach has been successfully used in a growing number of empirical studies: from astronomy to protein analysis, from image processing to finance, as well as in many other fields; see, e.g., Green et al. (2019); Kovacev-Nikolic et al. (2016); Bonis et al. (2016); Gidea and Katz (2018) and Wang et al. (2015). Theoretical understanding of relationships between statistical properties of different topological features and statistical properties of noisy data sets is the area of active ongoing research which could facilitate a broader application of TDA in many real-world complex systems; see Adler et al. (2010, 2014), Bobrowski and Mukherjee (2015), Chazal et al. (2016, 2018), and Vejdemo-Johansson and Mukherjee (2018) among others.

In this paper, we establish the functional dependency of the mean value of L^p-norms of persistence landscapes on a uniform scaling of the underlying multivariate distribution. Furthermore, we demonstrate that the average value of L^p-norms is decreasing when the covariance between the components of a m-dimensional random vector variable X is increasing. To illustrate the complex

1

dependency of these topological features on changes in the distribution, we conduct numerical experiments utilizing bi-variate distributions with known statistical properties. The high sensitivity of the mean of L^p-norms to transformation of the underlying multivariate distribution, reflected by changes in the covariance matrix Σ, has never been discussed in the literature and is significant from the practitioner's point of view.

Following the work of Gidea and Katz (2018), we apply TDA to noisy time series observed on financial markets. Sudden regime changes – financial crashes – are often characterized by increased market variability and covariance between broad market indices. This behavior is similar to critical transitions in many natural systems caused by endogenous forces; see Scheffer (2009) and Scheffer et al. (2009) and references therein. Here, we use the sliding window technique to quantify the temporal changes in total variations and co-variances between the components of X as well as evolution of L^1-norms of persistence landscapes derived from the daily log-returns of portfolio-like mixtures of major equity indices on European and US markets.

Our theoretical results and numeric experiments facilitate interpretation of the puzzling behavior of these topological features at the *inception phase* of the global meltdown caused by the COVID-19 pandemic. Contrary to early phases of the technology crash (March 2000) and the global financial crisis (2008-2009), despite the spike in variance of underlying time series, we do not observe any growth of L^p-norms in March 2020. We conclude that a strong synchronization between all elements of the global financial system nullifies persistence of transient loops that are usually present in point clouds representing noisy data sets.

In Section 2 we provide the background allowing readers to acquire a general understanding of the TDA methodology. In Section 3 we consider dependency of the mean of functional *p*-norms of persistence landscapes on changes in the underlying multivariate distribution. This consideration illuminates our empirical results, presented in Section 4. We offer conclusions in Section 5.

2. Background

Calculation of persistence homology (PH) is at the core of TDA; see Carlsson (2009), Carlsson and Zomorodian (2005), Edelsbrunner and Harer (2009). Informally, it is based on the computation of persistence of r-dimensional homologies, e.g., connected components ($r = 0$), loops ($r = 1$), cavities ($r = 2$), etc., at a wide range of scales.

2.1 Computation of Persistence Homology

The computational algorithm of PH involves construction of simplicial complexes, ordered with respect to the scaling parameter ϵ. Similarly to a network graph, the basic objects (simplices) include vertices (0-dimensional simplices) and edges (1-dimensional simplices). However, a simplicial complex includes higher dimensional simplices as well. For example, a filled triangle is a 2-dimensional simplex determined by its vertices (0-simplices), edges (1-simplices) and a face (2-simplex). If we eliminate the face, this construction becomes a 1-dimensional complex, corresponding to a loop, $r = 1$ homology. In that way, the simplicial complex provides a proxy for the shape of a point cloud, which represents a discrete multidimensional data set embedded in $\mathbb{R}^m$.

Extensive work has been done considering several types of complexes; see Carlsson (2009) and Ghrist (2008), among others. For instance, the popular Vietoris-Rips construction scheme contrives complexes by setting $\epsilon > 0$ for an edge of a r-simplex, $\sigma = [p_0, \ldots, p_r]$, iff the distance $d(p_i, p_j) \leq \epsilon$ for all i, j. The basic principle underlying computation of PH relies on the fact that alteration of the scale parameter ϵ results in modification of a simplicial complex. Consequently, homological attributes characterizing a simplicial complex are intrinsically dependent on ϵ. As the scaling parameter changes, some homologies appear while others disappear. Each homology is assigned a *birth* and a *death* value, and the difference between these two values represents its persistence. The output of this filtration procedure is captured in a concise form by a persistence (Rips) diagram. Coordinates

2

of each point on the Rips diagram represent the birth value (x-coordinate) and the death value (y-coordinate) of a r-dimensional hole. As a result, an arbitrary finite multidimensional data set is projected via the Rips filtration onto the two-dimensional persistence diagram. For a detailed exposition we refer to Edelsbrunner and Harer (2009).

2.2 The persistence landscape

Another PH-based summarization instrument is the persistence landscape, introduced by Bubenik (2015). It consists of a sequence of piece-wise linear functions defined in the re-scaled birth-death coordinates of the corresponding Rips diagram. The key advantage of persistence landscapes vs. persistence diagrams is related to their embedding into a Banach space. Consequently, one can apply standard tools of functional analysis and statistics to compute, e.g., their means, variances and p-norms.

Here, we formally define the persistence landscape and provide some of its properties. For a more detailed and broader exposition, we refer to Bubenik (2015, 2019).

Definition 1 *Let $\mathcal{D} = \{(b_i, d_i)\}_{i \in I}$ be a persistence diagram. For each birth-death point (b_i, d_i) in $\mathcal{D}$, we define a piecewise linear continuous function:*

$$f_{(b_i, d_i)}(x) = \begin{cases} x - b_i & \text{if } b_i < x \leq \frac{b_i + d_i}{2} \\ -x + d_i & \text{if } \frac{b_i + d_i}{2} < x < d_i \\ 0 & \text{otherwise.} \end{cases}$$

Then, the function $\Lambda \colon \mathbb{N} \times \mathbb{R} \longrightarrow \mathbb{R}$ given by

$$\Lambda(k, x) = \text{kmax}\{f_{(b_i, d_i)}(x)\}_{i \in I}, \tag{1}$$

is called the persistence landscape function associated to the persistence diagram $\mathcal{D}$, where kmax *denotes the k-th largest value of a set.*

Remark 2 *By definition, if $k > |I|$, then the value of kmax is zero, where I denotes the index set of $\mathcal{D}$ and $|I|$ denotes its cardinality.*

Alternatively, the persistence landscape may also be viewed as a sequence of functions $\lambda_1, \lambda_2, \ldots \colon \mathbb{R} \to \mathbb{R}$, where $\lambda_k(x) = \Lambda(k, x)$ is called the *k-th persistence landscape function* of $\mathcal{D}$. Each function $\lambda_k(x)$ is piecewise linear with slope either 0, 1, or -1. The critical points of λ_k are those values of x at which the slope changes. The set of critical points of the persistence landscape Λ is the union of the sets of critical points of the functions λ_k.

An overview of the persistence landscape associated to a given data set is illustrated in Figure 1, top right. For a persistence landscape Λ derived from a persistence diagram $\mathcal{D}$, the different k-th persistence landscape functions λ_k are clearly positive and by definition satisfy $\lambda_k(x) \geq \lambda_{k+1}(x)$ for all x.

Persistence landscapes can be understood in terms of elements of a Banach space. Recall that, for a given measure space $(\mathcal{S}, \mathcal{A}, \mu)$ and a function $f \colon \mathcal{S} \to \mathbb{R}$ defined μ-almost everywhere, for $1 \leq p < \infty$, the functional L^p-*norms* are defined as

$$\|f\|_p = \left(\int_{\mathcal{S}} |f|^p \, d\mu \right)^{1/p}.$$

Moreover, for $1 \leq p < \infty$, we have the Banach space

$$\mathcal{L}^p(\mathcal{S}) = \{f \colon \mathcal{S} \to \mathbb{R} \mid \|f\|_p < \infty\},$$

and define $L^p(\mathcal{S}) = \mathcal{L}^p(\mathcal{S})/\sim$, where $f \sim g$ if $\|f - g\|_p = 0$. Hence, we can define the norm of persistence landscape, as follows:
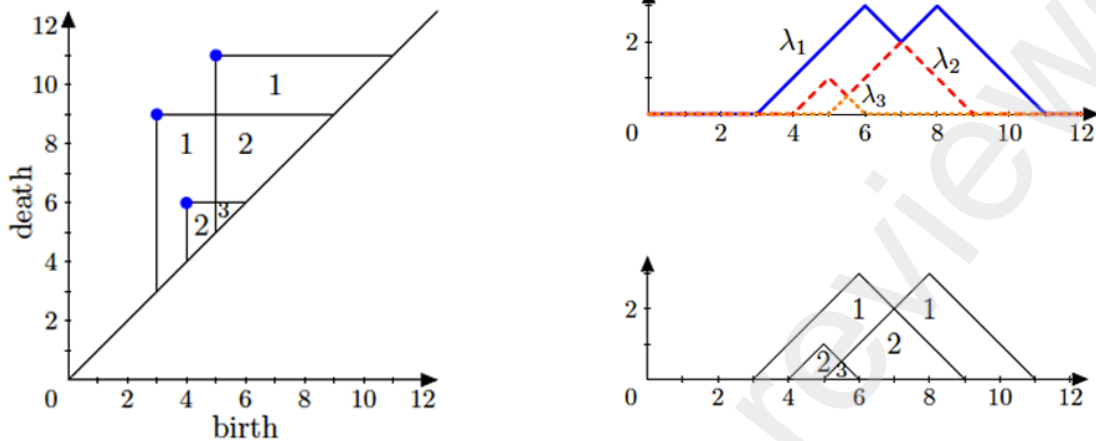
3

Figure 1: On the right, two visualizations of the persistence landscape derived from the persistence diagram on the left. The top-right figure, shows the different k-persistence landscape functions for $k = 1, 2, 3$. The figure is adapted from Bubenik (2015).

Definition 3 *Let $\Lambda\colon \mathbb{N} \times \mathbb{R} \to \mathbb{R}$ be a persistence landscape function. Suppose that on $\mathbb{N} \times \mathbb{R}$ we use the product of the counting measure on $\mathbb{N}$ and the Lebesgue measure on $\mathbb{R}$. Then, for $1 \leq p < \infty$, we define*

$$\|\Lambda\|_p = \sum_{k=1}^{\infty} \|\lambda_k\|_p \tag{2}$$

where $\lambda_k(t) = \Lambda(k, t)$, and $\|\lambda_k\|_p$ denotes the standard L^p-norm of λ_k.

Thus, we can endow the space of persistent landscapes with the norm (2) and the set of persistence landscapes becomes a subset of the Banach space $L^p(\mathbb{N} \times \mathbb{R})$.

3. Average values of p-norms of persistence landscapes

Our goal throughout this section is to characterize the dependency of the mean of L^p-norms of persistence landscapes on changes in the underlying multivariate distribution. For a given probability space $(\Omega, \mathcal{F}, P)$, let X denote a m-dimensional $(m > 1)$ random vector with a multivariate distribution function F. We can generate a point cloud that represents a sample of X of size N and compute the corresponding persistence landscape Λ. In this sense, we can interpret the persistence landscape as a Banach space valued random variable $\Lambda(\omega)\colon \Omega \to L^p(\mathbb{N} \times \mathbb{R})$, where the image space is endowed with the norm given in (2).

The analysis of the present paper are restricted to $1-$homologies, that is, loops. The number of loops in an N-point data set can be roughly bounded by the sum

$$L_1(N) := \sum_{i=4}^{i=N} \binom{N}{i} \leq 2^N, \tag{3}$$

where we consider the total number of loops to be bounded by grouping each different cluster of $i = 4, 5, \ldots, N$ points that can determine a 1D homology; notice that in the Vietoris-Rips construction scheme the minimum size of these clusters is 4.

4

First, we are going to derive the functional dependency of the average value of L^p-norms of persistence landscapes on a scaling factor of F. Throughout the exposition below, we assume regular conditions on F and rely on the Vietoris-Rips construction scheme.

3.1 Some properties of p-norms

For a given persistence landscape Λ, its L^p-norm is defined as $\|\Lambda\|_p = \sum_{k=1}^{\infty} \|\lambda_k\|_p$, where λ_k is the k-th persistence landscape function of Λ. It would seem obvious that for a finite data set, this sum is finite. Indeed, a finite data set describes a persistence diagram $\mathcal{D}^*$ composed of finitely many points (also with finite multiplicity). Thus, there exists t such that $\lambda_k = 0$ for all $k \geq t$ and $\|\lambda_k\|_p = 0$ for all p, hence $\sum_{k=1}^{\infty} \|\lambda_k\|_p < \infty$ (see Eq. (1)). To simplify our exposition, we denote $\mathcal{D}^* := \mathcal{D} \setminus \Delta$ to be the *undecorated persistence diagram* associated to a given point cloud, where Δ denotes the set of points on the diagonal of the persistence diagram $\mathcal{D}$; see Chazal et al. (2016).

Proposition 4 *Let $X \colon \Omega \to \mathbb{R}^m$ denote a random vector with a generic distribution function $F(\mu, \Sigma)$, where μ is the mean vector and Σ the variance-covariance matrix. Let $X^1, \ldots, X^N$ be independent identical copies of X such that $\mathbb{X} := (X^1, \ldots, X^N)$ describes an N-point data set in $\mathbb{R}^m$. Assume μ and Σ are finite. Then,*

$$\mathrm{E}(\|\Lambda\|_p) \leq L_1(N) \cdot N \cdot Tr(\Sigma)^{\frac{p+1}{2p}},$$

where $\|\Lambda\|_p$ denotes the p-norm of the persistence landscape Λ introduced in (2), and $L_1(N)$ is the bound of the number of loops introduced in (3).

Proof Let $\mathcal{D}^*$ denote the undecorated persistence diagram with corresponding persistence landscape Λ. From previous arguments we have that $|\mathcal{D}^*| < \infty$. Thus, we can define $\epsilon_b = \min_s\{b_s\}$ and $\epsilon_d = \max_l\{d_l\}$, to be the minimum parameter at which a generator is born and the maximum parameter at which a generator dies, respectively. Now, for each $x \in (\epsilon_b, \epsilon_d)$, we consider the piecewise linear function

$$\tilde{f}(x) := f_{(\epsilon_b, \epsilon_d)}(x) = \begin{cases} x - \epsilon_b & \text{if } x \in (\epsilon_b, \frac{\epsilon_b + \epsilon_d}{2}] \\ -x + \epsilon_d & \text{if } x \in (\frac{\epsilon_b + \epsilon_d}{2}, \epsilon_d). \end{cases}$$

Recall that in Section 2 we defined the function f_{p_i} for every $p_i \in \mathcal{D}^*$ so that $\lambda_k = \mathrm{kmax}\{f_{p_i}\}_{i \in I}$. To prove our statement, we can bound λ_1, and since $\lambda_k \geq \lambda_{k+1}$, this is enough to bound $\sum_{k \geq 1} \|\lambda_k\|_p$. It is easily deduced that $\mathrm{dom}(\lambda_1) = \mathrm{dom}(\tilde{f}) =: S \subset \mathbb{R}$, since $\lambda_1 = \max\{f_{p_i}\}_{i \in I}$. Moreover, $\tilde{f}(x) \geq \lambda_1(x) \geq 0$ for all $x \in S$ and taking integrals on both sides, for any $p \geq 1$, we have

$$\int_S (\tilde{f}(x))^p \, dx \geq \int_S (\lambda_{(}x))^p \, dx.$$

Thus,

$$\|\tilde{f}\|_p^p = \frac{(\epsilon_d - \epsilon_b)^{p+1}}{2^p(p+1)} \geq \|\lambda_1\|_p^p.$$

Now we can write the following chain of inequalities:

$$\begin{aligned} \|\lambda_1\|_p^p &\leq \frac{(\epsilon_d - \epsilon_b)^{p+1}}{2^p(p+1)} \overset{(*)}{\leq} \max_{1 \leq i \leq N} (d(X^i, \mu)^p) \\ &= \max_{1 \leq i \leq N} \Big(\sum_{j=1}^d (X_j^i - \mu_j)^2 \Big)^{\frac{p+1}{2}}. \end{aligned}$$

5

Inequality (*) comes directly from the fact that

$$(\epsilon_d - \epsilon_b) \leq 2 \cdot \max_{1 \leq i \leq N}(d(X^i, \mu)) = 2r.$$

In other words, the distance between two points of $\mathbf{X}$ inside the disc $B(\mu, r) = \{x \in \mathbf{R}^d \mid d(x, \mu) \leq r\}$ will never surpass $2r$. Taking expected values,

$$\mathrm{E}[\|\lambda_1\|_p] \leq \mathrm{E}\Big[\max_{1 \leq i \leq N}\Big(\sum_{j=1}^{d}(X_j^i - \mu_j)^2\Big)^{\frac{p+1}{2p}}\Big]$$

$$\leq \mathrm{E}\Big[\sum_{i=1}^{N}\Big(\sum_{j=1}^{d}(X_j^i - \mu_j)^2\Big)^{\frac{p+1}{2p}}\Big].$$

From here, setting $\alpha := 2p/(p+1)$ one can prove using Jensen's inequality that the following holds:

$$\mathrm{E}\Big[\sum_{i=1}^{N}\Big(\sum_{j=1}^{d}(X_j^i - \mu_j)^2\Big)^{\frac{1}{\alpha}}\Big] \leq N\Big(\sum_{j=1}^{d}\mathrm{E}(X_j - \mu_j)^2\Big)^{\frac{1}{\alpha}}$$

$$\leq N \cdot (\mathrm{Tr}\Sigma)^{\frac{1}{\alpha}}.$$

Notice that $\sum_{k \geq 1}\|\lambda_k\|$ has a finite number of non zero terms bounded by $L_1(N)$. Hence

$$\mathrm{E}(\|\Lambda\|_p) = \mathrm{E}\Big(\sum_{i=1}^{L_1(N)}\|\lambda_i\|_p\Big) = \sum_{i=1}^{L_1(N)}\mathrm{E}(\|\lambda_i\|_p)$$

$$\overset{(**)}{\leq} L_1(N) \cdot \mathrm{E}(\|\lambda_1\|_p) \leq L_1(N) \cdot N \cdot \mathrm{tr}(\Sigma)^{\frac{p+1}{2p}}.$$

Inequality (**) is due to the property $\lambda_k \geq \lambda_{k+1}$ mentioned in Section 2.2. ∎

Another useful property results from the analysis of changing values of L^p-norms in response to a uniform (isotropic) scaling of a point cloud by a certain factor $h > 0$.

Proposition 5 *Let $\mathbb{X}$ denote a point cloud in $\mathbb{R}^m$ with associated persistence landscape $\Lambda_{\mathbb{X}}$, and $H \colon \mathbb{R}^m \to \mathbb{R}^m$ be a homothetic transformation of the metric space such that $H(\mathbb{X}) = h \cdot \mathbb{X}$. Then,*

$$\|\Lambda_{H(\mathbb{X})}\|_p = h^{\frac{p+1}{p}} \cdot \|\Lambda_{\mathbb{X}}\|_p,$$

where $\|\Lambda_{H(\mathbb{X})}\|_p$ denotes the L^p-norm of the persistence landscape associated to the uniformly scaled point cloud $H(\mathbb{X})$.

Proof Consider $\mathcal{D}, \mathcal{D}'$ to be the corresponding persistence diagrams for $\mathbb{X}, H(\mathbb{X})$, respectively. Then, for every $p_i \in \mathcal{D}$ becomes $hp_i \in \mathcal{D}'$. Indeed, for any r-simplex $\sigma = [p_0, \ldots, p_r]$ in the Vietoris-Rips complex attached to $\mathbb{X}$ at parameter α, it has to be so that $d(p_i, p_j) \leq \alpha$ for all i, j. So, when we apply H we have that $\sigma = [p_0, \ldots, p_r]$ becomes $\sigma' = [hp_0, \ldots, hp_r]$ and if σ is formed at α, then σ' is formed at $h \cdot \alpha$. Thus, $|\mathcal{D}| = |\mathcal{D}'|$ and every point in $\mathcal{D}$ is scaled by h in $\mathcal{D}'$. Furthermore, for every f_{p_i} (mentioned in Section 2.2) associated to the persistence landscape of $\mathbb{X}$, we have that $f_{p_i} \mapsto f_{hp_i}$ and so, $\Lambda_{\mathbb{X}}$ becomes $h\Lambda_{\mathbb{X}}$. Recall that we denote $\Lambda_{\mathbb{X}}(k, x) = \lambda_k(x)$ to be the k-th persistence landscape function of $\Lambda_{\mathbb{X}}$, so we can write $\Lambda_{H(\mathbb{X})}(k, x) = h\Lambda_{\mathbb{X}}(k, x) = \lambda'_k(x)$. Note that it is enough to prove that

$$\|\lambda'_k\|_p^p = h^{p+1}\|\lambda_k(x)\|_p^p.$$

6

If λ'_k is null out of $(a', b') \subseteq (0, \infty)$, we have that λ_k is null out of (a, b) where $a' = ha$ and $b' = hb$. Then,

$$\int_{a'}^{b'} |\lambda'_k(y)|^p dy = h^{p+1} \int_a^b |\lambda_k(x)|^p dx.$$

∎

Now, we can characterize the dependency of the mean of L^p-norms of persistence landscapes on a uniform scaling of the underlying multivariate distribution.

Theorem 6 *Let $X \colon \Omega \to \mathbb{R}^m$ denote a random vector with the generic distribution function $F(\mu, \Sigma)$, where μ is the mean vector and Σ is the covariance matrix. Let $X^1, \ldots, X^N$ be independent identical copies of X such that $\mathbb{X} := (X^1, \ldots, X^N)$ describes a random N-point data set in $\mathbb{R}^m$. Assume μ and Σ are finite and that for a certain scaling factor h, $hF(\mu, \Sigma) = F(h\mu, h^2\Sigma)$. Then,*

$$\mathrm{E}(\|\Lambda_{h^2\Sigma}(\omega)\|_p) = h^{\frac{p+1}{p}} \cdot \mathrm{E}(\|\Lambda_\Sigma(\omega)\|_p), \tag{4}$$

where $\Lambda_{h^2\Sigma}$, Λ_Σ denote the persistence landscapes for random point clouds with corresponding covariance matrices $h^2\Sigma$ and Σ, respectively.

Proof Let H denote a homothetic transformation such that $H \colon \mathbb{R}^m \to \mathbb{R}^m$, so $\mathbf{X} \mapsto h \cdot \mathbf{X}$. From the assumptions made, we can safely deduce that for every point $X^i \in \mathbf{X}$ we have $H(X^i) = h \cdot X^i \sim F(h\mu, h^2\Sigma)$. From this point forward we will be using the following notation terms indistinguishably: $\Lambda_{\mathbf{X}}(\omega) = \Lambda_\Sigma(\omega)$ and $\Lambda_{h\cdot\mathbf{X}}(\omega) = \Lambda_{h^2\Sigma}(\omega)$.

Let $\mathbb{X}_1, \ldots, \mathbb{X}_n$ denote independent identically distributed copies of $\mathbb{X}$, and let $\Lambda_{\mathbb{X}}^1, \ldots, \Lambda_{\mathbb{X}}^n$ be the corresponding persistence landscapes. Proposition 5 implies that

$$\frac{1}{n} \sum_{i=1}^n \|\Lambda_{h^2\Sigma}^i(\omega)\|_p = h^{\frac{p+1}{p}} \frac{1}{n} \sum_{i=1}^n \|\Lambda_\Sigma^i(\omega)\|_p, \tag{5}$$

since for every landscape Λ_Σ^i we have the corresponding *scaled* persistence landscape $\Lambda_{h^2\Sigma}^i$. From Proposition 4 we have that the expected values $\mathrm{E}(\|\Lambda_\Sigma^i\|_p)$, $\mathrm{E}(\|\Lambda_{h^2\Sigma}^i\|_p)$ are finite. Hence, by the Strong Law of Large Numbers, we obtain the following two almost surely convergences:

$$\frac{1}{n} \sum_{i=1}^n \|\Lambda_{h^2\Sigma}^i(\omega)\|_p \longrightarrow \mathrm{E}(\|\Lambda_{h^2\Sigma}\|_p)$$

$$h^{\frac{p+1}{p}} \frac{1}{n} \sum_{i=1}^n \|\Lambda_\Sigma^i(\omega)\|_p \longrightarrow h^{\frac{p+1}{p}} \cdot \mathrm{E}(\|\Lambda_\Sigma\|_p)$$

Therefore, taking limits in (5), we obtain

$$\mathrm{E}(\|\Lambda_{h^2\Sigma}\|_p) = h^{\frac{p+1}{p}} \cdot \mathrm{E}(\|\Lambda_\Sigma\|_p),$$

as we claimed. ∎

The generality of Theorem 6, which does not depend on dimensionality or the size of a data set is rooted in the known *geometrical* property of a homothetic transformation. A uniform scaling alters the geometry of the point cloud, while its topology remains unchanged. However, this transformation scales the persistence of each homology in the cloud, which is reflected by the dependency of L^p-norms of persistence landscapes on the scaling factor h; see Proposition 5. In the statistical context, it results in Eq.4.

7

Theorem 7 *Let (X, Y) be a 2D-random vector sampled from a generic bi-variate distribution $\mathcal{D}(\mu, \Sigma)$, where $\mu = (\mu_x, \mu_y)$ denotes the 2D finite mean-vector and Σ is the 2x2 finite variance-covariance matrix:*

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_y\sigma_x & \sigma_y^2 \end{pmatrix}. \tag{6}$$

Here, $\rho \in (-1, 1)$ is the correlation coefficient. Assume $\sigma_x \geq \sigma_y$. Let θ_1 and θ_2 be the eigenvalues of the matrix Σ. Since Σ is symmetric and positive semi-definite, we have $\theta_1 \geq \theta_2 \geq 0$. For a fixed number of points N, consider the associated point cloud $\mathbb{X}$ and the corresponding persistence landscape Λ. Then, we have

$$\mathrm{E}[\|\Lambda\|_1] \leq L_1(N)^2 \cdot N \cdot \theta_2. \tag{7}$$

Proof Without loss of generality, using Principal Component Analysis, we can write $\mathbb{X}$ as a centered point cloud such that the points (X_i, Y_i), for $i = 1, \ldots, N$, are sampled from a bi-variate generic distribution $D(0, \Theta)$ where

$$\Theta = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix}.$$

The new covariance structure allows to interpret the X and Y axis as the axis of maximum and minimum variability, respectively. Now, we have that both variables are uncorrelated with the variance-covariance structure Θ, while the underlying topology of $\mathbb{X}$ remains unchanged. Furthermore, we can encapsulate $\mathbb{X}$ into a rectangle of basis $2\max_i |X_i|$ and height $2\max_i |Y_i|$ with $\max_i |Y_i| \leq \max_i |X_i|$.

Analogously we can encapsulate any subset of q points $(4 \leq q \leq N)$ of the point-cloud into a rectangle of basis α and height β where α is the maximum distance in the set of points converted in $x-$axis and β the corresponding orthogonal maximum $y-$distance. Of course, $\beta \leq \alpha$. The persistence of the 1D homology of this set of points is given by $\epsilon_d - \epsilon_b \leq \beta$. Indeed, if we denote $\phi_\alpha(x) := \sqrt{\alpha^2 + x^2}$ for $x \in [0, \alpha]$ we have

$$\epsilon_d - \epsilon_b \leq \phi_\alpha(\beta) - \phi_\alpha(0) = \phi_\alpha'(z)\beta$$

with $0 \leq z \leq \beta$. But $\phi_\alpha'(z) = \frac{z}{\sqrt{\alpha^2 + z^2}} \leq 1$.

Finally it is clear that for any set of points of the point cloud $\beta \leq 2\max_i |Y_i|$.

Recall now that for each point $(\epsilon_{b_i}, \epsilon_{d_i})$ in the persistence diagram, we have an associated *triangle* in the first persistence landscape function with the area given by $(\epsilon_{d_i} - \epsilon_{b_i})^2/4$; see (Bubenik, 2015). Hence, we can bound $\|\lambda_1\|_1$ of $\mathbb{X}$ as follows:

$$\|\lambda_1\|_1 \leq L_1(N) \cdot \max_i \frac{(\epsilon_{d_i} - \epsilon_{b_i})^2}{4}$$
$$\leq L_1(N) \cdot \max_i |Y_i|^2 \leq L_1(N) \cdot \Sigma_{i=1}^N |Y_i|^2.$$

From our proof of Proposition 4, we have that $\|\Lambda\|_1 \leq L_1(N) \cdot \|\lambda_1\|_1$, so taking expectations in the inequality above results in Inequality (7). ∎

Remark 8 *Note that from Theorem 7 it is easy to see that as $\rho \to \pm 1$, the expected value $\mathrm{E}[\|\Lambda\|_1] \to 0$. Indeed, the eigenvalues of Σ are:*

$$\theta_1 = \frac{\sigma_x^2 + \sigma_y^2}{2} + \sqrt{\left(\frac{\sigma_x^2 + \sigma_y^2}{2}\right)^2 - (1 - \rho^2)\sigma_x^2\sigma_y^2},$$
$$\theta_2 = \frac{\sigma_x^2 + \sigma_y^2}{2} - \sqrt{\left(\frac{\sigma_x^2 + \sigma_y^2}{2}\right)^2 - (1 - \rho^2)\sigma_x^2\sigma_y^2}.$$

Thus, if $\rho = 0$, the eigenvalues are σ_x^2, σ_y^2, and if $\rho = \pm 1$, we obtain $\sigma_x^2 + \sigma_y^2$ and 0.

8

The boundary on $E[\|\Lambda\|_1]$ that follows from Theorem 7 is rough. Nonetheless, its dependency on θ_2 points to suppression of PH in systems with strong covariance.

Remark 9 *If we have a point cloud $\mathbb{X}$ of an $m-$dimensional random vector in $\mathbb{R}^m$ we have analogously a covariance matrix which is a diagonal matrix composed by $\theta_1 \geq \cdots \geq \theta_m \geq 0$. And similarly to the 2-dimensional case, the persistence of loops is controlled by θ_2 and formula (7) is still valid.*

Remark 10 *Theorem 7 can be extended to $r-$dimensional holes for any $r \geq 0$. For any $r \geq 0$, the persistence of the $r-$holes is controlled by the $(r+1)-$th eigenvalue, provided $0 \leq r \leq m-1$. That is, connected components are bounded by $L_0(N)^2 N\theta_1$, loops are bounded by $L_1(N)^2 N\theta_2$, cavities are bounded by $L_2(N)^2 N\theta_3$ and so on. Here,*

$$L_r(N) = \sum_{i=2^{r+1}}^{i=N} \binom{N}{i},$$

of course, provided $N \geq 2^{r+1}$.

3.2 Numeric experiments

To illustrate the dependency described by Theorem 6, we construct $2D$ point clouds of two independent random variables by repetitive sampling of $N = 50$ points from joint distributions formed by two independent univariate Normal and (separately) by two independent univariate Gamma distributions. Notice that Theorem 6 does not require components of the random vector variable to be independent. The design of our numeric experiment is determined by simplicity of interpretation of its outcome. For each realization of the generated data set we associate the corresponding Rips filtration, compute the persistence diagram, the corresponding persistence landscape and its L^p-norm. We use the R-package TDA (Fasy et al. (2015)) for all related computations and calculate the mean-values of L^p-norms at the end of each simulation, which consists of 1,000 resamplings. We run 10 simulations per type of the joint distribution, sequentially increasing variances σ_i^2 $(i = 1, 2)$ of individual univariate distributions. For the joint Normal distribution, we set the initial values $\sigma_1 = \sigma_2 = \sigma = 1$ and increase them from 1 to 10 in equal increments. For the joint Gamma distribution, we chose the shape parameter $k1 = k2 = 2$. Recall that the univariate Gamma distribution is determined by the shape parameter k and the rate parameter β with $\mu = \frac{k}{\beta}$ and $\sigma^2 = \frac{k}{\beta^2}$. Therefore, for simulations with Gamma distributions, we set the initial values $\sigma_1 = \sigma_2 = \sigma = \sqrt{2}$ and again multiply them by h growing from 1 to 10 in each simulation. Figure 2 shows the outcome of these experiments. Notably, sampling from the joint Normal distribution (equivalently, from the uncorrelated bi-variate Normal distribution) results in practically the same plots; not shown here for brevity.

By design of our experiments, a uniform scaling of the underlying 2D distributions by h is equivalent to scaling of the total variation, σ^2, of these distributions by h^2. Thus, plots on Fig.2 confirm and illustrate the forecast of Theorem 6. Moreover, the Monte Carlo simulations clearly demonstrate that the predicted functional dependency, Eq.(3), is not sensitive to the shape of underlying distributions.

To illustrate the influence of changing covariance on average values of L^p-norms, we use sampling from the bi-variate Normal distribution. The choice of this model is motivated by the simplicity of its correlation structure. We use the mvrnorm function from the R-package MASS (Ripley and Venables (2002)) to sample $N = 50$ random data points from the bi-variate Normal distribution $\mathcal{N}(\mu, \Sigma)$, where $\mu = (\mu_x, \mu_y)$ is the 2D mean-vector and Σ is the 2x2 covariance matrix that has the structure given in Eq.(6). These experiments are inspired by the theoretical results obtained in Theorem 7 and Remark 8 regarding the dependency of the average value of L^1 norms on the correlation coefficient ρ of a bi-variate distribution. In the first experiment, we run 10 Monte Carlo simulations with $\sigma_x = \sigma_y = 1$ and sequentially increase ρ from 9.9% to 99.0% in 10 equal increments.
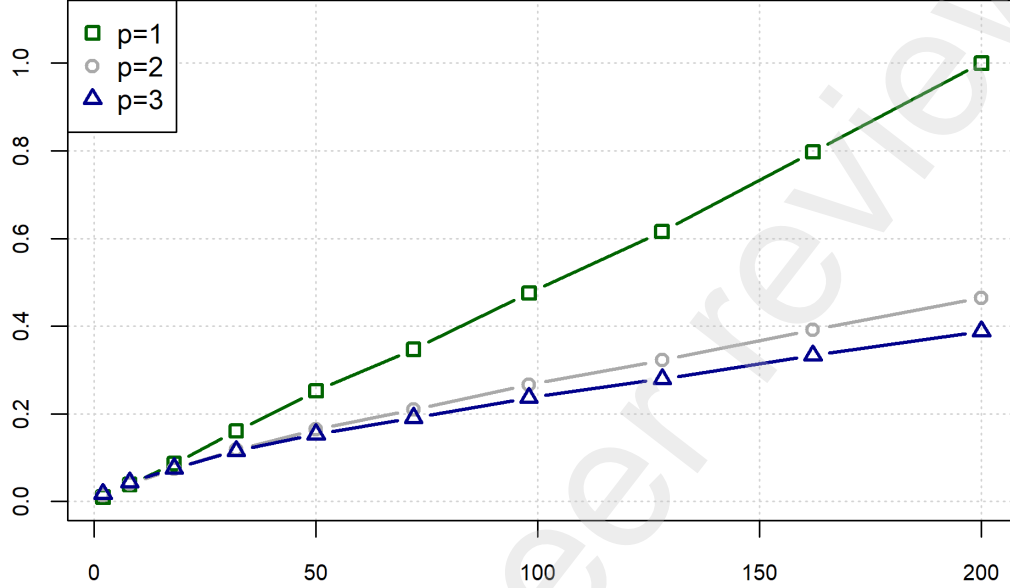
9

Figure 2: Plots of the dependency of average L^p-norms on the total variation of the joint distribution formed by two independent univariate Gamma distributions. Different markers denote the mean-values of L^p-norms (divided by the maximum of $E[L^1]$), computed at different values of p. Values of σ^2 are on the x-axis; see text for details.

In this case, we find that the average values of L^1-norms converge with the number of repetitions towards a smoothly decreasing function of the correlation coefficient. In the second experiment, we keep $\sigma_x = 1$ and increase σ_y from 0.3 to 3, while simultaneously increasing ρ from 9.9% to 99.0% in 10 equal increments. Finally, we run 10 simulations with $\sigma_x = \sigma_y$, both increasing from 0.3 to 3, simultaneously with ρ increasing from 9.9% to 99.0% in 10 equal increments. As expected, the latter two experiments demonstrate strong non-monotonic dependencies of the average values of L^1-norm on ρ^2. For simplicity, in all experiments we set $\mu = 0$. Figure 3 shows the outcome of these simulations.

Notice that the standardized generalized variance (SGV) of the bi-variate Normal distribution is proportional to the area of the respective confidence ellipse, S; see, e.g., Tong (1990). That is

$$SGV = \sqrt{|\Sigma|} = \sigma_x \sigma_y \sqrt{1 - \rho^2} \propto S \tag{8}$$

For $\rho = 0$ and $\sigma_x = \sigma_y = \sigma$, the bi-variate Normal distribution has the shape of a symmetric bell-curve in three dimensions with $SGV = \sigma^2$. It follows from Theorem 6 and numeric experiments that in this case, the mean value of L^1-norms is linearly *increasing* with σ^2; see Fig.2. One could reasonably assume that for weakly correlated random variables this dependency continues to be valid, which could qualitatively explain the initial linear increase of the average value of L^1-norms
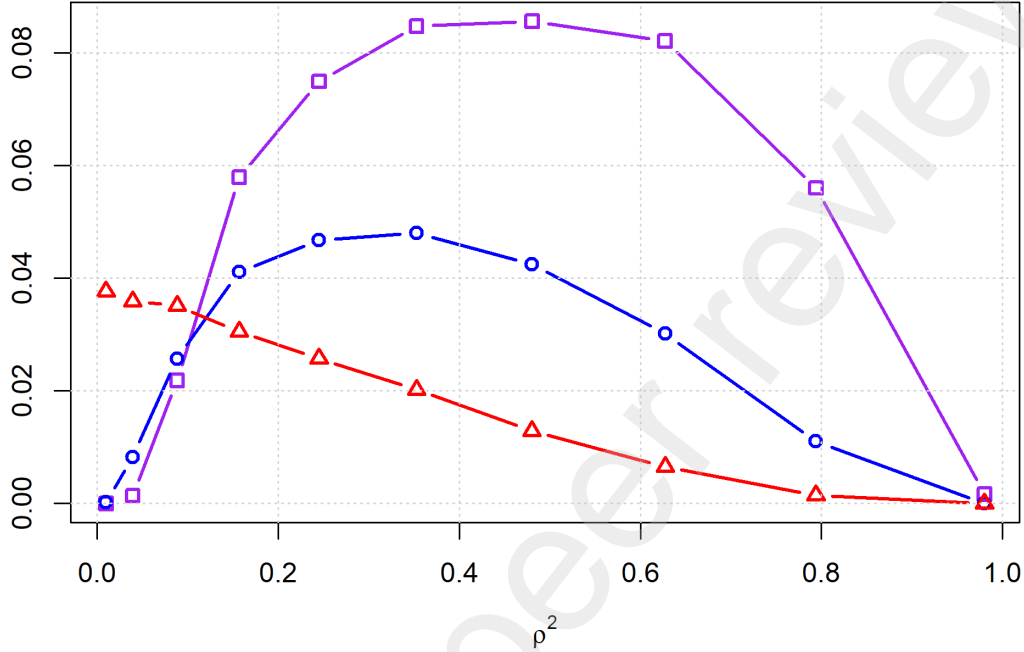
10

Figure 3: Dependency of the mean value of L^1-norms (y-axis) on parameters of Σ (10,000 repetitions). The line marked by triangles represents the case with $\sigma_x = \sigma_y = 1$. The line marked by circles represents the outcome of simulations with $\sigma_x = 1$ and σ_y growing from 0.3 to 3 in 10 equal increments simultaneously with ρ. The line marked by squares is derived for $\sigma_x = \sigma_y$ growing from 0.3 to 3 simultaneously with ρ in 10 equal increments.

with the rise of variability. On the other hand, for a small non-zero ρ, SGV and the area of the confidence ellipse are linearly *decreasing* with ρ^2; see Eq.(7). This could qualitatively describe the linear decrease of $E[L^1]$, when $\sigma_x = \sigma_y = \sigma$; see Fig.3. With a growing correlation, contours of isodensity stretch more in the direction of the main diagonal. As $\rho^2 \to 1$, contours degenerate into a straight line segment having zero persistence of $k = 1$ homologies and, therefore, the average value of p-norms of persistence landscapes approaches zero. Generally, this behavior follows from Theorem 7, see Eq.(6) and Remark 8, when ρ^2 approaches one. Figure 3 illustrates a complex dependency of the mean value of L^1-norms on variance and covariance of the bi-variate Normal distribution. In particular, it demonstrates that an increase of covariance between random variables can completely suppress the rise of norms of persistence landscapes due to a growing variability of the underlying multivariate distribution.

4. Empirical results

To be applied, the TDA-based dimensionality reduction requires an embedding of the data set into some, e.g., Euclidean, metric space. Any portfolio-like 'mixture' of financial time series naturally

11

defines the dimensionality of such a space, whereas the size of the sliding window determines the size of a point cloud; see Gidea and Katz (2018) for details. The time-resolved TDA methodology was successfully applied in multiple studies ranging from signal processing to financial forecasting; see Seversky et al. (2016); Berwald et al. (2014, 2015); Donato et al. (2006); Gidea (2017); Gidea et al. (2020); Harer and Perea (2015); Khasawneh and Munch (2016, 2018); Kramár et al. (2016); Vejdemo-Johansson et al. (2013); Maletić et al. (2016); Perea et al. (2015), and recent reviews in Ravishanker and Chen (2019) and Dłotko et al. (2019). We use Yahoo Finance API to obtain time
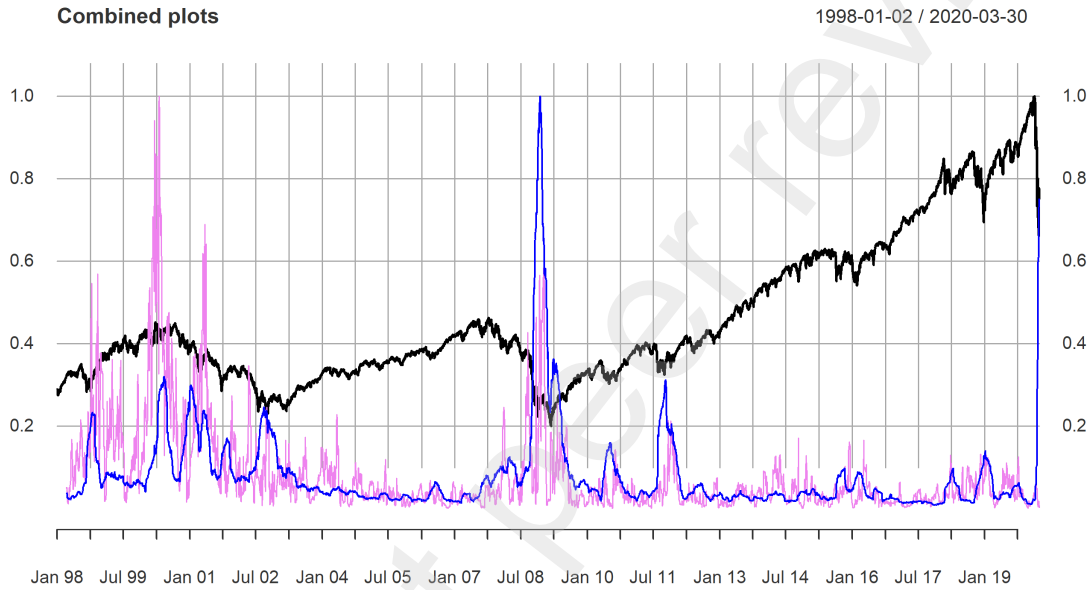

Figure 4: (Color online) Combined time series of S&P 500 (black line), L^1-norm (purple line) and estimated variability (blue line) of the mixture of daily log-returns of four indices; see text for details.

series of four broad US equity indices – DJI (DJI), S&P 500 (GSPC), NASDAQ (IXIC), and Russell 2000 (RUT) as well as four European indices FTSE 100 (FTSE), DAX 30 (GDAXI), CAC 40 (FCHI), IBEX 35 (IBEX) between January 1, 1998 and March 30, 2020. To synchronize trading dates on different European markets, we drop 3.9% of the data points. We arrange the mean-stationary time series formed by the daily log-returns of these indices, $ln(p_{i,t+1}/p_{i,t})$, where i identifies the index and t determines the trading day. We group them into two portfolio-like 'mixtures' - one with four US indices and another with four European indices. To derive the temporal changes in the relevant variance-covariance matrices as well as in the persistence of $1D$ loops, we use a sliding window of $n = 50$ trading days with the rolling step of one trading day to obtain two ordered in time sets of $4D$ point clouds, X_n^i, where $i = 1, 2$ determines the regional portfolio. The algorithm computes the corresponding Rips filtration $R(X_n^i, \epsilon)$, $\epsilon > 0$, the persistence diagram $\mathcal{D}_1(X_n^i)$, the respective persistent landscape $\Lambda(X_n^i)$, and its L^1-norm, $\|\Lambda(X_n^i)\|_1$, per point cloud.

Figure 4 shows the time series of S&P 500 (in black) as well as the time series of L^1-norm (in purple) and the total variation (in blue), derived for portfolio of U.S. indices. Notably, the time series of these features derived for portfolio of European indices with the rolling window of 50 trading days are practically the same and not shown here for brevity. The striking differences in the time series of L^1-norms at the *early stages* of the technology crash of 2000, the global financial crisis of

12

2008-2009, and the meltdown caused by COVID-19 pandemic (March, 2020), are clearly visible on these figures. First, notice that spikes in variability during the two prior systemic market crashes were always accompanied by peaks in the time series of L^1-norms. Remarkably, this behavior is not observed at the beginning of the latest global crash. Another noticeable observation is related to an exceptionally strong cross-correlation between indices, which was 95% or above for most of them in March 2020. These findings reflect the fundamentally different character of the global financial crisis triggered by a strong exogenous shock - COVID-19 pandemic.

5. Concluding remarks

TDA offers methods allowing for dimensionality reduction as well as means to infer 'shapes' of noisy data sets. Thereby, it provides a new perspective on many problems in the domain of complex systems. This paper combines the rigorous theoretical analysis with numerical experiments and the empirical study of a real-world complex multivariate system. We uncover the dependency of the expected value of L^p-norms of persistence landscapes on changes in the underlying multivariate probability distributions. This strong and complex dependency is rooted in general geometric and topological properties of these distributions. Our results are intuitively clear and should be taken into account in a growing number of applications utilizing L^p-norms of persistence landscapes as a key topological feature.

We demonstrate that if covariance between time series under study is low, the expected value of L^p-norms always moves in the same direction as the total variation in a system. This behavior was observed at early and later stages of the technology crash of 2000 and the global financial crisis of 2008 - 2009. Notice, however, that these market meltdowns were due to *endogenous* economic forces. On the other hand, the simultaneous drop of all market indices, triggered by the *exogenous* COVID-19 shock, in March 2020, leads to an unusually strong correlation between indices. The latter translates into a 'line flattening' of the respective point clouds, which nullifies persistence of all k-dimensional homologies ($k > 0$). Consequently, despite the strong spike in the total variation, we do not find any rise of the time series of L^p-norms till the end of March 2020.

References

Robert J. Adler, Omer Bobrowski, Matthew S. Borman, Eliran Subag and Shmuel Weinberger. Persistent homology for random fields and complexes. In *Borrowing strength: theory powering applications*, volume 6, pages 124–143. Inst. Math. Statist., Beachwood, OH, 2010.

Robert J. Adler, Omer Bobrowski and Shmuel Weinberger. Crackle: the homology of noise. *Discrete Comput. Geom.*, 52(4):680–704, 2014.

Jesse Berwald, Marian Gidea. Critical transitions in a model of a genetic regulatory system. *Math. Biosci. Eng.*, 11(4):723–740, 2014.

Jesse Berwald, Marian Gidea, Mikael Vejdemo-Johansson. Automatic recognition and tagging of topologically different regimes in dynamical systems. *Discontinuity, Nonlinearity and Complexity*, 3(4):413–426, 2015.

Omer Bobrowski and Sayan Mukherjee. The topology of probability distributions on manifolds. *Probability Theory and Related Fields*, 161(3-4):651–686, 2015.

13

Thomas Bonis, Maks Ovsjanikov, Steve Oudot and Frédéric Chazal. Persistence-based pooling for shape pose recognition. In *6th International Workshop on Computational Topology in Image Context (CTIC 2016)*, pages 19–29. Springer-Verlag Berlin, Heidelberg, 2016.

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1):77–102, 2015.

Peter Bubenik (2020) The Persistence Landscape and Some of Its Properties. In: Baas N., Carlsson G., Quick G., Szymik M., Thaule M. (eds) Topological Data Analysis. Abel Symposia, vol 15. Springer, Cham. https://doi.org/10.1007/978-3-030-43408-3

Gunar Carlsson and Afra Zomorodian. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.

Gunar Carlsson. Topology and data. *Bull. Amer. Math. Soc.*, 46(2):255–308, 2009.

Frédéric Chazal, Vin de Silva, Marc Glisse and Steve Oudot. *The Structure and Stability of Persistence Modules.* SpringerBriefs in Mathematics, Springer International Publishing, 2016.

Frédéric Chazal, Brittany Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo and Larry Wasserman. Robust topological inference: distance to a measure and kernel distance. *J. Mach. Learn. Res.*, 18(159):1–40, 2018.

Paweł Dłotko, Wanling Qiu and Simon Rudkin Cyclicality. Cyclicality, Periodicity and the topology of time series. arXiv:1905.12118 [math.AT], 2019.

Irene Donato, Matteo Gori, Marco Pettini, Giovanni Petri, Sarah De Nigris, Roberto Franzosi, and Francesco Vaccarino. Persistent homology analysis of phase transitions. *Phys. Rev. E*, 93(5):052138, 2016.

Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction.* Amer. Math. Soc., Providence, RI, 2009.

Robert Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc.*, 45(1):61–75, 2008.

Marian Gidea. Topology data analysis of critical transitions in financial networks. In *3rd International Winter School and Conference on Network Science*, pages 47–59. Springer Proceedings in Complexity. Springer, Cham, 2017.

Marian Gidea and Yuri Katz. Topological data analysis of financial time series: landscapes of crashes. *Physica A*, 491(1):820–834, 2018.

Marian Gidea, Daniel Goldsmith, Yuri Katz, Pablo Roldan and Yonah Shmalo. Topological recognition of critical transitions in time series of cryptocurrencies. *Physica A*, 548(1):123843, 2020.

Sheridan B. Green, Abby Mintz, Xin Xu and Jessi Cisewski-Kehe. Topology of our cosmology with persistent homology. *Astronomy and Computing*, 27(1):34–52, 2019.

John Harer, Jose A. Perea. Sliding windows and persistence: an application of topological methods to signal analysis. *J. Found. Comput. Math.*, 15(3):799–838, 2015.

Firas A. Khasawneh, Elizabeth Munch. Chatter detection in turning using persistent homology. *Mechanical Systems and Signal Processing*, 70-71(1):527–541, 2016.

Firas A. Khasawneh, Elizabeth Munch. Topological data analysis for true step detection in periodic piecewise constant signals. *Proc. R. Soc. A*, 474(2218):20180027, 2018.

14

Violeta Kovacev-Nikolic, Peter Bubenik, Dragan Nikolic, and Giseon Heo. Using persistent homology and dynamical distances to analyze protein binding. *Stat. Appl. Genet. Mol. Biol.*, 15(1):19–38, 2016.

Miroslav Kramár, Rachel Levanger, Jeffrey Tithof, Balachandra Suri, Mu Xu, Mark Paul, Michael F. Schatz, Konstantin Mischaikow. Analysis of Kolmogorov flow and Rayleigh-Benard convection using persistent homology. *Physica D*, 334(1): 82–98, 2016.

Slobodan Maletić, Yi Zhao and Milan Rajković. Persistent topological features of dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(5):53105, 2016.

Jose A. Perea, Anastasia Deckard, Steve B. Haase and John Harer. SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data. *BMC Bioinformatics*, 16(1):257, 2015.

Nalini Ravishanker and Renjie Chen. Topological data analysis (TDA) for time series. arXiv:1909.10604 [stat.AP], 2019.

Brian Ripley and Bill Venables. *Modern Applied Statistics with S*. Springer, New York, NY, 2002. Latest software available at https://CRAN.R-project.org/package=MASS.

Marten Scheffer. *Critical Transitions in Nature and Society*. Princeton Studies in Complexity. Princeton University Press, Princeton, NJ, 2009.

Marten Scheffer, Jordi Bascompte, William A.Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. van Nes, Max Rietkerk and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, 2009.

Lee M. Seversky, Shelby Davis and Matthew Berger. On time-series topological data analysis: New data and opportunities. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1014–1022, 2016.

Yung L. Tong. *The Multivariate Normal Distribution*. Springer Series in Statistics. Springer-Verlag New York, NY, 1990.

Mikael Vejdemo-Johansson, Gurjeet Singh, Muthu Alagappan, John Carlsson and Gunnar Carlsson. Extracting insights from the shape of complex data using topology. *Scientific Reports*, 3(1):1236, 2013.

Brittany T. Fasy, Jisu Kim, Fabrizio Lecci, Clément Maria, David L. Millman, and Vincent Rouvreau. Statistical tools for topological data analysis. Software available at https://CRAN.R-project.org/package=TDA, 2015.

Mikael Vejdemo-Johansson and Sayan Mukherjee. Multiple testing with persistent homology. arXiv:1812.06491 [cs.CG], 2018.

Yuan Wang, Hernando Ombao, and Moo K Chung. Topological epilepsy seizure detection in electroencephalographic signals. In *Proceedings. IEEE International Symposium On Biomedical Imaging*, pages 351–354, 2015.