Persistent homology for random fields and complexes

Robert J. Adler[*,†,1]   **Omer Bobrowski**[*,†,‡,2]   **Matthew S. Borman**[†,3]
Eliran Subag[*,†,4]   **and Shmuel Weinberger**[†,§,5]

Electrical Engineering, Technion – Israel Institute of Technology Department of Mathematics, University of Chicago

Abstract: We discuss and review recent developments in the area of applied algebraic topology, such as persistent homology and barcodes. In particular, we discuss how these are related to understanding more about manifold learning from random point cloud data, the algebraic structure of simplicial complexes determined by random vertices and, in most detail, the algebraic topology of the excursion sets of random fields.

1. Introduction

Over the last few years there has been a very interesting and rather exciting development in what is reputedly one of the most esoteric areas of pure mathematics: algebraic topology. Some of the practitioners of this subject are, to a considerable extent, looking out beyond the inner beauty of their subject and seeing if they can apply it to problems in the 'real world', that is to problems outside the realm of pure mathematics. As a result 'applied algebraic topology' is no longer an oxymoron, and although it is true that at this point sophisticated applications are still few and far between, there is a growing feeling that the gap between theory and practice is closing. We shall give more specific references below, but a very lively discussion of this trend can be found in Rob Ghrist's review [23], book in progress [24] and website on a project on sensor topology for minimal planning [25]. Gunnar Carlsson's webpage [11], which describes a large Stanford TDA (topological data analyis) project, and a DARPA webpage [16] describing a broad based project,

[1]Electrical Engineering, Technion, Haifa, Israel 32000, e-mail: `robert@ee.technion.ac.il`, url: `webee.technion.ac.il/people/adler`
[2]Electrical Engineering, Technion, Haifa, Israel 32000, e-mail: `bober@tx.technion.ac.il`, url: `webee.technion.ac.il/people/bober/`
[3]Mathematics, University of Chicago 5734 S. University Ave, Chicago, IL 60637, e-mail: `borman@math.uchicago.edu`, url: `http://www.math.uchicago.edu/∼borman/`
[4]Electrical Engineering, Technion, Haifa, Israel 32000, e-mail: `selirans@t2.technion.ac.il`
[5]Mathematics, University of Chicago 5734 S. University Ave, Chicago, IL 60637, e-mail: `shmuel@math.uchicago.edu`, url: `http://www.math.uchicago.edu/∼shmuel/`

also help explain the reasons why so many people have been so attracted to this direction.

These ideas are not totally new. For example, the brain imaging community has been using random field modelling and topological properties of these fields for quite some time. For example, people like Karl Friston, a leading figure in medical imaging, have been talking about the notion of 'topological inference' for a while (cf. the website [22]) based in a large part on the work of the late Keith Worsley. What is new, however, is the coordinated attack of a goodly number of high powered mathematicians on applications.

The aim of the current paper is to describe some of the new ideas that have arisen in applied algebraic topology and, given the interests of the authors, exploit some of them in the setting of random fields, i.e. of random processes defined over spaces of dimension greater than one. There are new results here, albeit without proofs. However, this paper is mainly review and exposition, with a strong bias in a particular direction, but written in a language which we hope will be accessible to the natural readers of this Festchrift who may (as did we until recently) find the language of even applied topologists somewhat unfamiliar. In the final analysis, if Larry will be happy with the final product, then we shall be happy as well.

The paper starts in Section 2 with a discussion of one of the central notions of applied algebraic topology, that of persistent homology and its graphical depiction via barcodes. This is done via examples rather than formal definitions, so it should be possible to understand the notion of persistent homology without actually knowing what a homology group is. (For those who do know about homology, more precise definitions of persistent homology and barcodes are given in the appendix of Section 6.) Also in Section 2 we discuss simplicial complexes as they arise in manifold learning as well as discussing the topology of random field excursion sets.

Section 3 has a brief discussion of persistence diagrams of excursion sets, based on simulations. (These have actually already been used elsewhere for the analysis of brain imaging data, see [15].) These data raise numerous challenges for statisticians and probabilists.

Section 4 introduces what seems at first to be a rather abstruse structure of 'Euler integration', but it is very quickly shown that not only is this a useful concept, but the key to solving a number of quite varied problems. This is the main section of the paper.

A very brief Section 5 points out that we should have also had more to say about random simplicial complexes, but didn't, and so points you to appropriate references. A brief technical appendix completes the paper in Section 6.

2. Persistent homology and barcodes

In this section we are going to give a very brief and sketchy introduction to some basic notions of algebraic topology. A concise, yet very clear introduction to the topics that concern us can be found in [9, 24], while [26, 37] are good examples of a thorough coverage of homology theory. Recent excellent and quite different reviews by Carlsson [10, 13], Edelsbrunner and Harer [20], and Ghrist [1, 17, 18, 23] give a broad exposition of the basics of persistent homology.

Algebraic topology focuses on studying topology by assigning algebraic, group theoretic, structures to topological spaces X. Thus, homology, cohomology and homotopy groups can be used to classify objects into classes of 'similar shape'. In this paper we shall focus on homology. If X is of dimension N, then it has $N+1$ homology

groups, each one of which is an abelian group. (We shall later take the coefficients from $\mathbb{Z}_2$, thereby making the groups vector spaces.) The zero-th homology $H_0(X)$ is generated by elements that represent connected components of X. For $k \geq 1$ the k-th homology group $H_k(X)$ is generated by elements representing k-dimensional 'loops' in X. The rank of $H_k(X)$, denoted by β_k, is called the k-th *Betti number*. For X compact and $k \geq 1$, β_k, measures the number of k-dimensional holes in X, while β_0 counts the number of connected components. The Euler characteristic, a central topological quantity and homotopy invariant, is then

$$(2.1) \qquad \chi(X) \; = \; \sum_{k=0}^{N} (-1)^k \beta_k.$$

To explain the idea of persistent homology, we shall work with two examples. The first is based on what is known as the 'Morse filtration' of excursion sets, the second on complexes formed from point sets.

2.1. Barcodes of excursion sets

Suppose that M is a nice space, that $f : M \to \mathbb{R}$ is smooth, and consider the excursion, or super-level, sets

$$(2.2) \qquad A_u \; \triangleq \; \{p \in M \, f(p) \in [u, \infty)\} \; \equiv \; f^{-1}([u, \infty)).$$

Note that if $u \geq v$ then $A_u \subseteq A_v$. Going from u to v, components of A_u may merge and new components may be born and possibly later merge with one another or with the components of A_u. Similarly, the topology of these components may change, as holes and other structures form and disappear. Following the topology of these sets, as a function of u, by following their homology, is an example of persistent homology. The term 'persistence' comes from the fact that as the level u changes there is no change in homology until reaching a level u which is a critical point of f; i.e. the topology of the excursion sets remains static, or 'persists', between the heights of critical points. This, of course, is the basic observation of Morse Theory, which links critical points to homology. However, the persistence of persistent homology goes further. For example, when two components merge, one treats the first of these to have appeared as if it is continuing its existence beyond the merge level.

A useful way to describe persistent homology is via the notion of barcodes. Assuming that $\dim(M) = N$, we also have, from the smoothness of f, that, if A_u is non-empty, then $\dim(A_u)$ will typically also be N. A barcode for the excursion sets of f is then a collection of $N + 1$ graphs, one for each collection of homology groups of common order. A bar in the k-th graph, starting at u_1 and ending at u_2 ($u_1 \geq u_2$) indicates the existence of a generator of $H_k(A_u)$ that appeared at level u_1 and disappeared at level u_2. An example is given in Figure 2.1, in which the function f is actually the realisation of a smooth random field on the unit square, an example to which we shall return later.

Figure 2.2 is even more impressive, since it shows a three dimensional example. Note that, as opposed to the 2-dimensional case, it is almost impossible to say anything about topology just by looking at the boxes with the excursion sets at the top of the figure, but there is a lot of immediate visual information available in the barcodes. This phenomenon becomes even more marked as the dimension N of the parameter space increases. While it may be impossible to imagine what a six dimensional excursion set looks like, it is easy to look at a barcode with six sets of bars for the six persistent homologies.
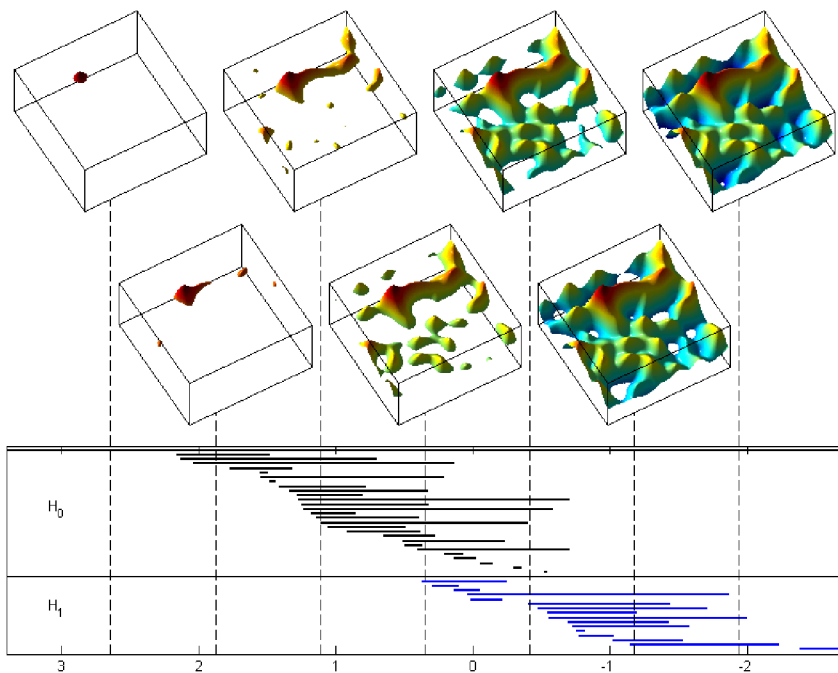
FIG 2.1. *Barcodes for the excursion sets of a function on $[0,1]^2$. The top seven boxes show the surfaces generated by a 2-dimensional random field above excursion sets A_u for different levels u. To determine the level for each figure, follow the vertical line down to the scale at the bottom of the barcode. As the vertical lines pass through the boxes labelled H_0 and H_1, the number of intersections with bars in the H_0 (H_1) box gives the number of connected components (resp. holes) in A_u. Thus, at $u \sim 1.9$, A_u has 4 connected components but no holes, while at $u \sim -1.2$, A_u has only 1 connected component, but 9 holes. The horizontal lengths of the bars indicate how long the different topological structures (generators of the homology groups) persist. Computation of the barcodes was carried out in Matlab using Plex (Persistent Homology Computations) from Stanford [12].*

2.2. Point sets and manifold learning

Consider the following situation. Let X be an unknown subset of $\mathbb{R}^d$ with finite Lebesgue measure and let $X_1, \ldots, X_n$ be n independent random samples uniformly distributed on X. We would like to study the homology of X using only these random points. When X is a manifold, this is typically referred to as *manifold learning*. In many cases we can find an ε for which the union of balls

$$(2.3) \qquad\qquad U \;=\; \bigcup_{i=1}^{n} B_\varepsilon(X_i)$$

is homotopy equivalent to X (and hence has the same homology). However, we do not know, a priori, what is the correct choice of ε. An example is given in Figure 2.3, in which X is a two-dimensional annulus. If ε is chosen to be too small then U is homotopy equivalent to the union of n distinct points (and hence contains no information on X). On the other hand, choosing ε to be too big gives us a U that is a large, contractible blob, which again tells us nothing about X. But, as with Goldilock's porridge, choosing ε 'just right', recovers an object topologically equivalent to the annulus. Persistent homology overcomes this sensitivity to the
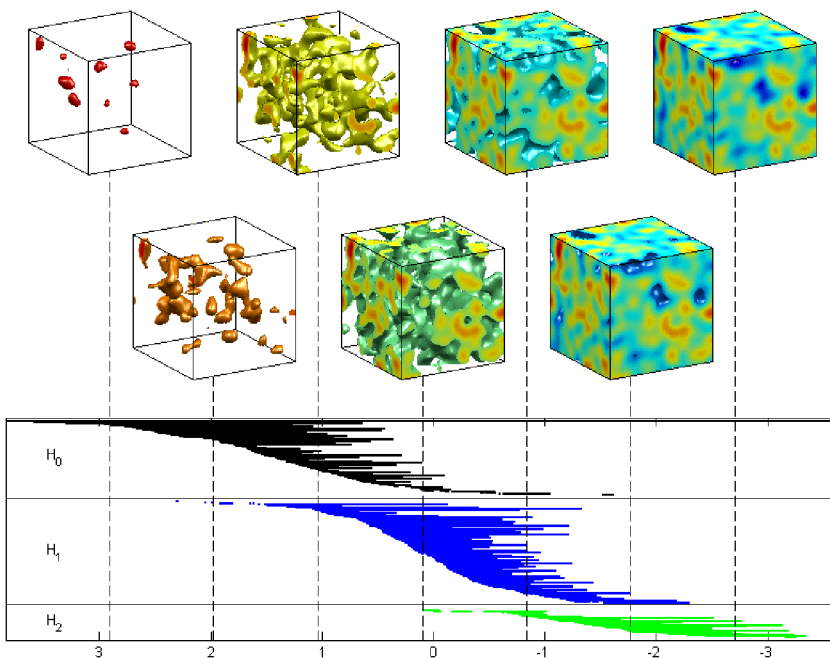
FIG 2.2. *Barcodes for the excursion sets of a 3-dimensional random field. The barcode diagram is to be read as for Figure 2.1, with two differences: The top 7 boxes now display the excursion sets themselves and the values of the field are colour coded. Furthermore, there are now three homology-groups/barcode-boxes, representing connected components, handles, and holes.*

choice of ε by considering a range of possible values of ε, much as we did with the levels of excursion sets in the previous example, but with the aim of learning about the topology of X from the barcodes. The key assumption is that homology elements that 'live longer' (or, *persist*) are more likely to represent homology elements of X, whereas the shorter ones are just 'noise'.

To describe this in a little more detail we need the notion of simplicial complexes.

2.3. Simplicial complexes

We are not going to give a definition of simplicial complexes here, but rather shall describe two classic ways to construct abstract simplicial complexes from a given set of points in a metric space.
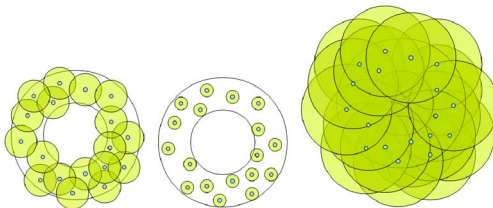
FIG 2.3. *Trying to capture the homology of an annulus (where $\beta_0 = 1$, $\beta_1 = 1$) from a union of balls of various radii around a random sample of n points from the annulus. A good choice of radius recovers the correct homology in the first case. If the radius chosen is too small, the union of balls has the same homology as n distinct points ($\beta_0 = n, \beta_1 = 0$). If the radius chosen is too big, the union is contractible ($\beta_0 = 1, \beta_1 = 0$).*

Definition 2.1 (The Čech complex). Let $\mathcal{P} = \{x_1, x_2, \ldots\}$ be a collection of points in a metric space X. Construct an abstract simplicial complex $C(\mathcal{P}, \varepsilon)$ in the following way:

1. The 0-simplices are the points in $\mathcal{P}$,
2. An n-simplex $[x_{i_0}, \ldots, x_{i_n}]$ is in $C(\mathcal{P}, \varepsilon)$ if $\bigcap_{k=0}^{n} B_\varepsilon(x_{i_k}) \neq \emptyset$,

where $B_\varepsilon(x)$ is the ball of radius ε around x. The complex $C(\mathcal{P}, \varepsilon)$ is called the Čech complex attached to $\mathcal{P}$ and ε.

Definition 2.2 (The Vietoris-Rips complex). Let $\mathcal{P} = \{x_1, x_2, \ldots\}$ a collection of points in a metric space X. Construct an abstract simplicial complex $R(\mathcal{P}, \varepsilon)$ in the following way:

1. The 0-simplices are the points in $\mathcal{P}$.
2. An n-simplex $[x_{i_0}, \ldots, x_{i_n}]$ is in $R(\mathcal{P}, \varepsilon)$ if $B_\varepsilon(x_{i_k}) \cap B_\varepsilon(x_{i_m}) \neq \emptyset$ for every $0 \leq k < m \leq n$.

The complex $R(\mathcal{P}, \varepsilon)$ is called the Rips complex attached to $\mathcal{P}$ and ε.

From these definitions it is obvious that $C(\mathcal{P}, \varepsilon) \subset R(\mathcal{P}, \varepsilon)$. In addition, it is proved in [17] that $R(\mathcal{P}, \varepsilon') \subset C(\mathcal{P}, \varepsilon)$ for $\varepsilon/\varepsilon' \geq \sqrt{2d/(d+1)}$. In other words, a Čech complex can be 'approximated' by Rips complexes. This fact is used in computational applications, since working with Rips complexes is much more efficient than with Čech complexes.

There are occasions when Rips and Čech complexes coincide, as is the case when X is Euclidean but the metric is the L^∞ rather than the more standard L^2 norm. In many statistical applications the choice of metric on X may be dictated by optimality considerations rather than 'natural' geometry.

The main importance of the Čech complex and its relevance to homology theory, is given in the next theorem.

Theorem 2.3 (The Nerve theorem). *Suppose that the intersections $\bigcap_{x \in \mathcal{P}'} B_\varepsilon(x)$ are either empty or contractible for any subset $\mathcal{P}'$ of $\mathcal{P}$. Then the Čech complex $C(\mathcal{P}, \varepsilon)$ is homotopy equivalent to $\bigcup_{x \in \mathcal{P}} B_\varepsilon(x)$. In particular, if X is a finite dimensional normed linear space, or a compact Riemannian manifold with convexity radius greater than ε, and if $\{B_\varepsilon(x)\}_{x \in \mathcal{P}}$ is a cover of the space X, then $C(\mathcal{P}, \varepsilon)$ is homotopy equivalent to X.*

The main consequence of the Nerve Theorem is that in order to study the homology of the *topological* space $\bigcup_{x \in \mathcal{P}} B_\varepsilon(x)$, we can study the homology of the *combinatorial* space $C(\mathcal{P}, \varepsilon)$. This fact can be useful in proving theoretical results, but its main contribution is to computational applications.

With these definitions behind us, Figure 2.4 gives a nice example of how barcodes describe the topology of an annulus ($\beta_0 = 1$, $\beta_1 = 1, \beta_2 = 0$) in $\mathbb{R}^2$, when 17 points are sampled from it and Rips complexes are computed for a range of ε.

3. Random field simulations

In this section we want to consider the persistent homology of random field excursion sets. In particular, we would like to understand something about the distributional properties of their barcodes.

The random fields behind the barcodes of Figures 2.1 and 2.2 were taken to be mean zero, Gaussian, over the parameter set $[0, 1]^2$ and with covariance function
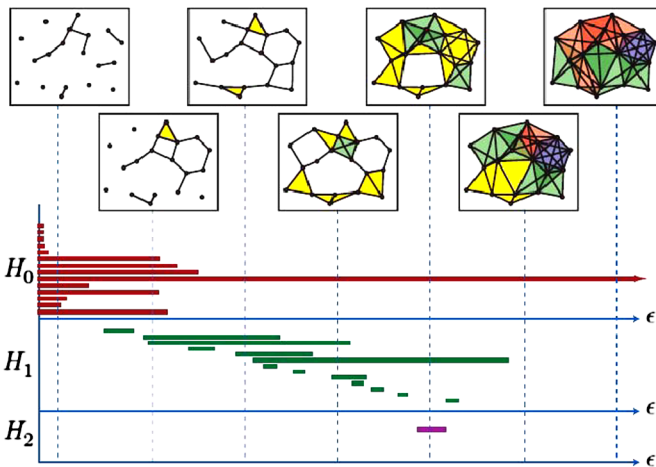
FIG 2.4. *The barcode of a Rips complex, taken from [23]. The points were sampled from an annulus in $\mathbb{R}^2$. We see that there is a single H_0 bar that persists forever. This bar represents the single connected component of the annulus. In H_1 we see a couple of dominant bars indicating that the sample space contains holes. The longest bar actually represents the real hole of the annulus. In H_2 there is nothing significant and indeed $\beta_2 = 0$ in this case.*

$R(p) = \exp(-\alpha\|p\|^2)$. This is a stationary, isotropic, and infinitely differentiable random field, and the starting test case for all theories. We took $\alpha = 100$.

We ran 10,000 simulations of this field, calculating 10,000 barcodes. In order to represent the data in a reasonable fashion, we used *persistence diagrams* rather than barcodes. To form a persistence diagram from the bars in H_k, one simply replaces each bar by a pair (x, y), where x is the level at which the bar begins and y the level at which it ends. Thus $x > y$ and the pair (x, y) lies in a half plane. In Figure 3.1 the corresponding persistence diagrams for the complete simulation data are shown for H_0 and H_1.

Additional information on the barcodes is given in Figure 3.2. What is shown there are the (marginal) distributions of the start and end points of the barcodes for H_0 and H_1 from the same simulation. A simple application of Morse theory, or, in this simple two dimensional setting, a little thought, leads to the realisation that the start points of the H_0 bars are all heights of local maxima of the field, while
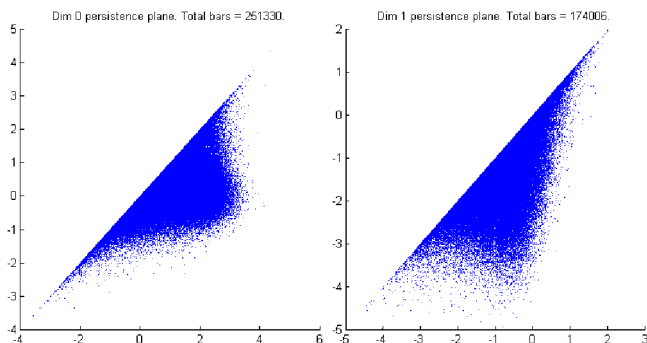

FIG 3.1. *Persistence diagrams for 10,000 simulations of an isotropic random field on the unit square. Note that the diagrams for H_0 and H_1 seem quite different.*
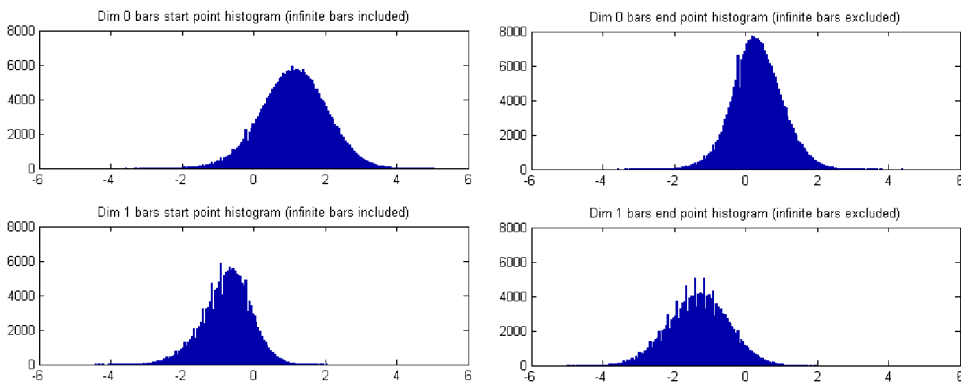
FIG 3.2. *Empirical distributions of start and end points of bars for the Gaussian field of Figure 3.1.*

the end points of the H_1 bars correspond to local minima. These distributions have been well studied (although their precise form is not known) in the general theory of Gaussian random fields. The remaining start and end points correspond to different types of saddle points of the random field. However, what differentiates between the end point of a H_0 bar and the start point of a H_1 bar is global geometry and is not determined by the local behaviour of the field.

It would be interesting to know more about the real distributions lying behind Figures 3.1 and 3.2, but at this point we know very little. An interesting aspect is the asymmetries between the start points of the H_0 bars (local maxima) and the end points of the H_1 bars (local minima), as well as between the two sets of saddle points. This is due to boundary effects and would disappear if the simulation had been carried out on a closed manifold.

There are some things that we do know, however, and we turn to them next. Firstly, however, we need to make a small digression.

4. Euler integration

Before we introduce the Euler integral, we need to define the Euler characteristic for noncompact spaces. For a compact space X, we already defined the Euler characteristic $\chi(X)$ in (2.1) as an alternating sum of Betti numbers; viz. as an alternating sum of ranks of the homology groups $H_k(X)$.

In this setting the Euler characteristic is a homotopy invariant and is additive in the sense that

$$(4.1) \qquad \chi(A \cup B) = \chi(A) + \chi(B) - \chi(A \cap B).$$

In extending the definition of the Euler characteristic to noncompact spaces, if one uses the definition of the Euler characteristic as the alternating sum of $\operatorname{rank} H_k(X)$, then additivity is lost (consider $[0,1] = [0,1) \cup \{1\}$). Therefore the definition of the Euler characteristic we shall use for noncompact spaces is

$$\chi(X) = \sum_k (-1)^k \operatorname{rank} H_k^{\mathrm{lf}}(X) = \sum_k (-1)^k (\# \text{ of open } k\text{-simplices in X}),$$

where $H_k^{\mathrm{lf}}(X)$ is called the locally finite homology (see [27, Chapter 3]). Since $H_k^{\mathrm{lf}}(X) = H_k(X)$ for compact spaces, this extends the definition of the Euler characteristic to noncompact spaces such that additivity is preserved, but we lose

homotopy invariance, although it is still a homeomorphism invariant. This Euler characteristic can also be computed by decomposing the space into a union of open k-simplices and points. For example, $\chi((0,1)) = -1$, $\chi([0,1)) = 0$ and $\chi([0,1]) = 1$.

4.1. The Euler integral

Since the Euler characteristic is an additive operator on sets (cf. (4.1)), it is tempting to consider χ as a measure and integrate with respect to it. The main problem in doing so is that χ is only finitely additive.

At first (cf. [38]), integration with respect to the Euler characteristic was defined for a small set of functions called *constructible functions* defined by

$$CF(X) = \left\{ h(x) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(x) \ \middle| \ n \in \mathbb{N}, a_k \in \mathbb{Z}, A_k \text{ is tame} \right\},$$

where 'tame' means having a finite Euler characteristic. For this set of functions we can define the Euler integral similarly to the Lebesgue integral. Let $h(x) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}(x)$ and define

$$\int_X h \, d\chi \ \triangleq \ \sum_{k=1}^{n} a_k \chi(A_k).$$

This integral has many nice properties, similar to those of the Lebesgue integral, such as linearity and a version of the Fubini theorem (cf. [24, 38]). However, due to the lack of countable additivity one cannot easily continue from here by approximation to integrate more general functions.

Nevertheless, in [3] two possible extensions were suggested for the Euler integral of real valued functions. We shall not go into details of the constructions here, but rather use one of the properties of these extensions to define what in [3] was called an upper Euler integral and which we, for simplicity, shall call an Euler integral. Thus, we define the Euler integral by

$$(4.2) \qquad \int_X f \, d\chi \ \triangleq \ \int_{u=0}^{\infty} \left[\chi(f > u) - \chi(f \le -u) \right] \, du,$$

where $\chi(f > u) \triangleq \chi\left(f^{-1}(u, \infty)\right)$ and $\chi(f \le -u) \triangleq \chi\left(f^{-1}[-u, \infty)\right)$. These integrals are defined for what are known as 'tame' functions. See [3] for details.

Unfortunately, these extensions of the Euler integral have many flaws, of which the most prominent one is the lack of additivity. For example, a simple computation shows that for $X = [0,1]$

$$\int_X x \, d\chi + \int_X (1-x) \, d\chi = 1 + 1 = 2 \neq 1 = \int_X 1 \, d\chi.$$

Nevertheless, these integrals still have interesting properties and some interesting applications. Here is one.

4.2. An application of the Euler integral

This application was suggested in [4]. Suppose that an unknown number of targets are located in a space X and each target α is represented by its support $U_\alpha \subset X$.

Suppose also that the space X is covered with sensors, each reporting only the number of targets it can sense, but with no ability to distinguish between targets. Let $h : X \to \mathbb{Z}$ be the *sensor field*, i.e.

$$h(x) = \# \{\text{targets activating the sensor located at } x\} .$$

The following theorem states how to combine the readings from all the sensors and get the exact number of targets.

Theorem 4.1 (Baryshnikov and Ghrist, [4]). *If all the target supports U_α satisfy $\chi(U_\alpha) = \gamma$ for some $\gamma \neq 0$, then*

$$\# \{\text{targets}\} = \frac{1}{\gamma} \int_X h \, d\chi.$$

Note that we do not need to assume anything about the targets other than they all have the same nonzero Euler characteristic. For example, we need not assume that they are all convex or even have the same number of connected components. On the other hand, the theorem assumes an ideal sensor field, in the sense that the entire (generally continuous) space X is covered with sensors which register only what happens at the point at which they are placed. In [3] more realizable models using the upper and lower Euler integrals are discussed.

Assume now that the readings from the sensors are contaminated by a Gaussian (or Gaussian related) noise $f(x)$. Under these conditions it can be proved that

$$\int_X (h + f) \, d\chi = \int_X h \, d\chi + \int_X f \, d\chi.$$

Denoting $s = \int_X h \, d\chi$ (deterministic signal), $n = \int_X f \, d\chi$ (noise) and $y = \int_X (h + f) \, d\chi$ (measurement), this is a classic signal plus noise problem (i.e. $y = s + n$). In particular, in order to estimate s from y, it would be nice, in view of Theorem 4.1, to be able to compute some distributional properties of the Euler integral of a Gaussian random field. We shall limit ourselves to computing the expectation and shall turn to this after a few words on the Gaussian kinematic formula.

4.3. The Gaussian kinematic formula

Suppose that M is an N-dimensional, C^2, Whitney stratified manifold satisfying some mild side conditions (cf. [2] for details) and D a similarly nice stratified sub-manifold of $\mathbb{R}^k$. Let $f = (f^1, \dots, f^k) : M \to \mathbb{R}^k$ be a vector valued random process, the components of which are independent, identically distributed, real valued, C^2, centered, unit variance, Gaussian processes. Using f, define a Riemannian metric on M by setting

$$(4.3) \qquad g_x(X, Y) \stackrel{\Delta}{=} \mathbb{E}\{(X f_x^i)\,(Y f_x^i)\},$$

for any i and for $X, Y \in T_x M$, the tangent space to M at $x \in M$, and use this to define the Lipschitz-Killing curvatures, $\mathcal{L}_j$, $j = 0, \dots, N$ on M. For example, if M is a manifold without boundary, then these are given by

$$(4.4) \qquad \mathcal{L}_j(M) = \frac{1}{(2\pi)^{(N-j)/2}((N-j)/2)!} \int_M \mathrm{Tr}^M(-R)^{(N-j)/2} \, \mathrm{Vol}_g,$$

when $N - j \geq 0$ is even, and 0 otherwise. Here Vol_g is the volume form of the Riemannian manifold (M, g), R is the curvature tensor and Tr^M the trace operator on the algebra of double forms on M. For simple Euclidean spaces, with various orderings and normalisations, the Lipschitz-Killing curvatures are also known as Quermassintegrales, Minkowski or Steiner functionals, integral curvatures, and intrinsic volumes. Note that $\mathcal{L}_N(M) \equiv \mathrm{Vol}_g(M)$ is the Riemannian volume of M and $\mathcal{L}_0(M) \equiv \chi(M)$ is its Euler characteristic.

The Gaussian kinematic formula (hereafter GKF) was due originally to Taylor in [35] (but for the form below see [2, 36]) and states that

$$(4.5) \quad \mathbb{E}\left\{\mathcal{L}_i\left(M \cap f^{-1}(D)\right)\right\} = \sum_{j=0}^{N-i} \begin{bmatrix} i+j \\ j \end{bmatrix} (2\pi)^{-j/2} \mathcal{L}_{i+j}(M) \mathcal{M}_j^\gamma(D).$$

The combinatorial coefficients here are the standard 'flag coefficients' of integral geometry, given by

$$\begin{bmatrix} n \\ j \end{bmatrix} = \binom{n}{j} \frac{\omega_n}{\omega_{n-j}\,\omega_j},$$

where ω_n is the volume of the unit ball in $\mathbb{R}^n$. The $\mathcal{M}_j^\gamma(D)$, known as the Gaussian Minkowski functionals of D, are determined via the tube expansion

$$(4.6) \quad \mathbb{P}\left\{f(x) \in \{y : d(y, D) \leq \rho\}\right\} = \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \mathcal{M}_j^\gamma(D),$$

where x is any point in M and d is the usual Euclidean distance from a point to a set.

One could devote a book to this formula and, indeed, such a book exists. So we shall refer you to [2] for all needed technical details.

We note only one pertinent fact, for immediate use. Taking $j = 0$ in (4.5) gives the expected Euler characteristic of excursion sets as a simple, closed form expression that can be readily calculated in many interesting cases. Again, see [2] for details.

4.4. The Euler integral of a Gaussian random field

Returning to the signal plus noise problem of Section 4.2, we can formulate the first step towards its solution.

Let M be a nice, tame, space. (The definition of 'tame' can be found in [2].) Let f be a random field. Here is a striking result, due to Bobrowski and Borman [6]:

Theorem 4.2. *Let M be an N-dimensional tame stratified space and let $f : M \to \mathbb{R}^k$ be a k-dimensional Gaussian random field satisfying the GKF conditions. Let $G : \mathbb{R}^k \to \mathbb{R}$ be piecewise C^2 and let $g = G \circ f$. Setting $D_u = G^{-1}(-\infty, u]$ and assuming that $\left| \int_{\mathbb{R}} \mathcal{M}_j(D_u) du \right| < \infty$, we have*

$$(4.7) \quad \mathbb{E}\left\{\int_M g\, d\chi\right\} = \chi(M)\mathbb{E}\{g\} - \sum_{j=1}^{N} (2\pi)^{-j/2} \mathcal{L}_j(M) \int_{\mathbb{R}} \mathcal{M}_j^\gamma(D_u)\, du,$$

where $\mathbb{E}\{g\} := \mathbb{E}\{g(t)\}$. (g(t) has constant mean.)

While, on the one hand, this is not a difficult result to prove, given the GKF and (4.2), it was completely unforseen until discovered and has a number of interesting and potentially deep implications.

The main difficulty in applying Theorem 4.2 lies in computing the Minkowski functionals $\mathcal{M}_j^\gamma(D_u)$. A simple example is given in the following case:

Theorem 4.3. *Let M be an N-dimensional tame stratified space and let $f : M \to \mathbb{R}$ be a real valued Gaussian random field satisfying the GKF conditions. Let $G : \mathbb{R} \to \mathbb{R}$ be piecewise C^2 and let $g = G \circ f$. Then*

$$\mathbb{E}\left\{\int_M g\, d\chi\right\} = \chi(M)\mathbb{E}\{g\} + \sum_{j=1}^{N}(-1)^j \mathcal{L}_j(M)\frac{\langle H_{j-1}, (\mathrm{sgn}(G'))^j G'\rangle}{(2\pi)^{j/2}}.$$

In the theorem, for $n \geq 0$ the nth Hermite polynomial H_n is defined by

$$H_n(x) = (-1)^n \varphi^{-1}(x)\frac{d^n}{dx^n}\varphi(x),$$

where φ is the standard Gaussian density, $e^{-x^2/2}/\sqrt{2\pi}$. The inner product is given by

(4.8) $$\langle f, g\rangle = \int_\mathbb{R} f(x)g(x)\varphi(x)\, dx,$$

and we use the convention, required below, that

$$H_{-1}(x) = \varphi^{-1}(x)\int_x^\infty \varphi(u)\, du.$$

In the case that the function G is strictly monotone, we have an even simpler form.

Corollary 4.4. *Let f be as in Theorem 4.3 and G be a strictly increasing function, then*

$$\mathbb{E}\left\{\int_M g\, d\chi\right\} = \sum_{j=0}^{N}(-1)^j \mathcal{L}_j(M)\frac{\langle H_j, G\rangle}{(2\pi)^{j/2}}.$$

If G is strictly decreasing then,

$$\mathbb{E}\left\{\int_M g\, d\chi\right\} = \sum_{j=0}^{N}\mathcal{L}_j(M)\frac{\langle H_j, G\rangle}{(2\pi)^{j/2}}.$$

Finally, taking G to be the identity function yields

Corollary 4.5. *Let f be as in Theorem 4.3, then*

$$\mathbb{E}\left\{\int_M f\, d\chi\right\} = -\frac{\mathcal{L}_1(M)}{\sqrt{2\pi}}.$$

Further details and further examples, including results for χ^2 and F random fields, can be found in [6].

4.5. Persistent homology of Gaussian excursion sets

We now return to excursion sets, which would seem to have been forgotten in all the discussion on Euler integrals. It turns out that this was not the case, but that Euler integrals and Theorem 4.3 contain a lot of information on the persistent homology of Gaussian excursion sets.

First, however, some notation and a definition: Suppose we have a barcode, which we shall denote by $\mathcal{B}$. Denote the individual bars in $\mathcal{B}$ by b, their lengths by $\ell(b)$, and the degree of the homology group to which belongs the generator that they represent by $\mu(b)$.

Definition 4.6. The *Euler characteristic* of a barcode $\mathcal{B}$ with no bars of infinite length is

$$\chi(\mathcal{B}) \triangleq \sum_{b \in \mathcal{B}} (-1)^{\mu(b)} \ell(b).$$

It turns out that this topological quantity can actually be written in terms of Euler integrals. For convenience, and adopting the prejudices of topologists rather than probabilists and statisticians, we shall consider the barcodes of incursion rather than excursion sets, or, equivalently, sub-level sets rather than super-level sets. That is, we replace excursion sets of (2.2) by

$$(4.9) \qquad \widetilde{A}_u \triangleq \{p \in M \ f(p) \in (-\infty, u]\} \equiv f^{-1}((-\infty, u]).$$

Then [6] showed that, if $\mathcal{B}(f, u)$ denotes the barcode of $\widetilde{A}_u$ for tame f,

$$(4.10) \qquad \chi(\mathcal{B}(f, f_{\max})) = f_{\max}\chi(M) - \int_M f \, d\chi.$$

Combining this with Theorem 4.2 yields

Theorem 4.7. *Let $f : M \to \mathbb{R}^k$ be a Gaussian random field satisfying the GKF conditions, $G \in C^2(\mathbb{R}^k, \mathbb{R})$ and $g = G \circ f$. Then*

$$\mathbb{E}\{\chi(\mathcal{B}(g, g_{\max}))\} = \chi(M)\left(\mathbb{E}\{g_{\max}\} - \mathbb{E}\{g\}\right) + \sum_{j=1}^{N} (2\pi)^{-j/2}\mathcal{L}_j(M) \int_{\mathbb{R}} \mathcal{M}_j(D_u)\, du.$$

If f is real, then

$$\mathbb{E}\{\chi(\mathcal{B}(f, a))\} = \chi(M)\left(\varphi(a) + a\Phi(a)\right) + \varphi(a)\sum_{j=1}^{N} (2\pi)^{-j/2}\mathcal{L}_j(M)H_{j-2}(-a),$$

for any a.

The way we have presented this result, as a 'natural' consequence of a reasonably 'straightforward' Theorem 4.2, substantially underplays its importance and novelty. It has a number of interesting corollaries, for which we send you to the original paper [6]. But its main contribution lies in its very existence, connecting, as it does, between probabilistic objects and their homological structure.

As one of our colleagues/teachers recently stated: "I can think of no two topics in mathematics further away from one another than probability and algebraic topology. There is probably no way to connect them." Yet here, in Theorem 4.7, is an elegant connection, one of the first of its kind.

5. Random geometric complexes

In Sections 2.2 and 2.3 we motivated the idea of persistent homology and barcodes with examples from manifold learning and random simplicial complexes. Despite this, we shall not go into detail, but shall rather describe some general issues and give you a few useful references to this area, which is also currently undergoing rapid development.

5.1. Manifold learning

We already mentioned manifold learning briefly in Section 2.2 via the example of trying to identify an annulus, or at least its homology, from a simplicial complex built over the sample points. The subject of manifold learning goes, obviously, well beyond such an example, and examples of algorithms for 'estimating' an underlying manifold from a finite sample abound in the statistics and computer science literatures. Very few of them, however, take an algebraic point of view, which is what we have stressed in this paper.

One contribution in the spirit of this paper is [30] by Niyogi, Smale, and Weinberger, who studied the problem of estimating smooth manifolds from finite samples. They showed that in sampling from a high dimensional manifold, if the sampling is dense enough then, with high probability, the set (2.3) deformation retracts to the manifold and so has the same homology. This implies that long persistence intervals, once one has enough sample points, are very likely to correctly compute the homology of a submanifold.

Of course, one of the most important issues in dealing with data is noise. In the setting of manifold learning this translates to the sample points possibly not coming from the submanifold that theoretically models the phenomenon because of experimental, measurement, or other error. In [31] the same authors treat this issue, as does [14] from a different and enlightening point of view.

In a complementary but related direction [8, 9] apply persistence techniques to the nonparametric study of functions on a given manifold.

5.2. Random complexes

We now return to the Čech and Rips complexes of Section 2.3.

To get a feeling for the phenomena that occur as we approximate a manifold by the union of balls, it is perhaps enlightening to consider the situation, for a fixed ε, of the evolution of the homology of the Čech complexes as the number of points, n, grows. The points themselves we assume are chosen uniformly, at random, on the manifold.

For n small, the balls do not intersect, but as n grows intersections begin to occur and small finite graphs appear. Assuming that ε is sufficiently small that all (geodesic) ε-balls have about the same volume, it is easy to compute the expected number of times a particular graph arises. This leads to complicated integrals, but investigating them leads to the belief that k-homology (for a Čech complex) is most likely to occur as a result of the occurrence of boundaries of $(k+1)$-simplices. That is, it requires $k+1$ points to be close to each other (at scale ε) but not to fill in. Aside from a constant factor, the probability that this happens should be $(\varepsilon^d)^k \binom{n}{k+1}$. For this probability to be non-trivial, one requires that n is $O(1/\varepsilon^{dk/(k+1)})$. In other words, it is for n of this size that one begins to see interesting k-homology.

As n continues to grow, there will be a point where the ε-balls cover a nontrivial percentage of the volume, a point at which phenomena related to percolation occur. Finally, there is a reversal when the ε-balls fill almost all of the manifold, all extra homology dies, and ultimately we obtain the correct calculation of homology.

To get a feeling for the end game, it is worthwhile to compute the expected Euler characteristic of the union of n ε-balls in, say a flat torus. For simplicity, as we are only giving a heuristic, let's avoid the complications of Euclidean metrics and consider an L^∞ metric on the torus. In that case, a straightforward inclusion exclusion argument (see [32] for this in a Poisson model in Euclidean space and [29] for the use of kinematic formulae to obtain the relevant formulae in the case of genuine round balls), together with a generating function argument, give the formula for $\mathbb{E}(\chi_{n,\varepsilon,d})$, where $\chi_{n,\varepsilon,d}$ is the Euler characteristic of the union of n ε-balls and d is the dimension of the torus, as follows:

Let $\tau = (2\varepsilon)^d$. Then

$$
\mathbb{E}\left\{\chi_{n,\varepsilon,d}\right\} = \begin{cases} n(1-(\tau)^{n-1}), & d = 1, \\ \dfrac{d}{d\tau}\left(\tau\mathbb{E}\left\{\chi_{n,\tau,d-1}\right\}\right), & d \geq 2. \end{cases}
$$

One thus obtains that the Euler characteristic is approximately 0 (for the last time and so implying coverage) when n is around $(1/\tau)\log(1/\tau)$ plus lower order terms. See [21] for more details. Interestingly, it follows from the work of [33] that the phase transition for the giant component to form, in the sense of random graph theory, is (asymptotically) at 2^{-d} times this number. That is, the computation of components seems to be correct. It also seems extremely likely that there are phase transitions at other multiples of this fundamental scale where the other homology groups are correctly computed.

Many more details of the phenomena for small n and the percolation range, the relevant central limit theorems for homology and some valuable information about persistence intervals in Rips and Čech situations can be found in recent papers of Kahle [28] and Bobrowski and Borman [7]. They combine probabilistic tools with Morse theory to give rigorous proofs of these phenomena. We mention also an early lecture by Diaconis, available on the web [19], which suggests the general outline of this picture, and a forthcoming paper [5] that also deals with some aspects of this problem in a metric-measure setting (relevant to situations where the distribution of points in nature does not follow the Riemannian uniform measure).

6. Technical appendix

As promised, we shall now be a little more formal and explain what persistent homology really is. For this, however, we shall need to assume that the reader has a basic working familiarity with the theory of simplicial homology. We shall also, for simplicity, take all homologies over the group $\mathbb{Z}_2$. This is the material of Section 6.1.

The second subsection of the appendix explains how to carry these definitions over to random complexes. To complete the Appendix we should have really added a section on how one turns the excursion sets of a continuous random field into a random filtered complex. This, as you will be able to guess, after reading the first two subsections, is done by discretizing the parameter space of the random field and then thresholding the field at various levels in order to obtain the simplices of the filtered complex. You can find details of this in the report [34].

6.1. Persistent Homology

We start by considering growing sequences of simplicial complexes that grow in the following manner.

Definition 6.1. A filtered simplicial complex is a sequence of simplicial complexes, $\mathcal{K} = \{K_j\}_{j \geq 0}$, such that

$$\widehat{C}_n\left(K_0\right) \subset \widehat{C}_n\left(K_1\right) \subset \widehat{C}_n\left(K_2\right) \subset \widehat{C}_n\left(K_3\right) \subset \cdots ,$$

for all $n \geq 0$, where $\widehat{C}_n(K)$ is the collection of all n-simplices in the complex K.

We say that a simplex σ enters the filtered complex $\mathcal{K}$ at the *entrance time i* if $\sigma \in K_i$ and $\sigma \notin K_j$ for all $j < i$. Occasionally, we shall use the term filtration instead of filtered complex.

The usual computation of the homology groups of the complexes K_j is done for each j at a time, and so does not allow for comparison of homologies between complexes. The idea of persistent homology is to take the filtration into account and so be able to describe how homological properties persist or disappear as k grows.

Denoting the n-cycles of a complex K by $Z_n(K)$ and the n-boundaries by $B_n(K)$, note that any cycle in $Z_n(K_j)$ also belongs to $Z_n(K_{j+1})$ and boundaries in $B_n(K_j)$ belong to $B_n(K_{j+1})$. This allows us to define the linear maps

$$i_*^{(n,j)} : H_n\left(K_j\right) \longrightarrow H_n\left(K_{j+1}\right),$$
$$\bar{z} = z + B_n\left(K_j\right) \longmapsto \bar{z} = z + B_n\left(K_{j+1}\right).$$

Definition 6.2. The p-persistent n-th homology group of K_j is defined by

$$H_n^p\left(K_j\right) = i_*^p\left(H_n\left(K_j\right)\right) \subset H_n\left(K_{j+p}\right),$$

where i_*^p denotes the composition

$$i_*^{(n,j+p-1)} \circ i_*^{(n,j+p-2)} \circ \cdots \circ i_*^{(n,j)}.$$

The non-zero elements of the persistent homology group are the images of n-cycles which exist at time j (i.e. belong to K_j) and which 'survive' until time $j+p$, in the sense that they are not nullified by becoming a boundary.

We now restrict the discussion to filtered complexes of finite type.

Definition 6.3. We say that a filtered simplicial complex $\mathcal{K} = \{K_j\}_{j \geq 0}$ is of finite type if, for all $j \geq 0$ and $n \geq 0$, $\widehat{C}_n\left(K_j\right)$ is finite and if there exists an index i such that $K_j = K_i$ for all $j \geq i$.

Now fix $n \geq 0$. Recall that we are working with $\mathbb{Z}_2$, so that the $H_n(K_j)$ are all vector spaces. Using algebraic arguments[1] it can be shown that for any filtered complex of a finite type it is possible to choose bases $\{c_1^j, c_2^j, \ldots, c_{m_j}^j\}$, of $H_n(K_j)$, for all $j \geq 0$, such that for any $1 \leq k \leq m_j$, $i_*(c_k^j) \in \{0, c_1^{j+1}, c_2^{j+1}, \ldots, c_{m_{j+1}}^{j+1}\}$ and $i_*(c_k^j) = i_*(c_{k'}^j)$ for $k \neq k'$ only if $i_*(c_k^j) = 0$.

Figure 6.1 shows an example of this relation between the bases for a certain filtered complex. The elements below each of the homologies form a basis of the homology. Note that we have written $H_k \equiv H_n(K_k)$ in order to save space.

---

[1]For details see [39].

$$
\begin{array}{ccccccc}
H_0 & H_1 & H_2 & H_3 & H_4 & H_5 & H_6 & \cdots \\
c_1^0 \overset{i_*}{\mapsto} & c_1^1 \overset{i_*}{\mapsto} & c_1^2 \overset{i_*}{\mapsto} & c_1^3 \overset{i_*}{\mapsto} & c_1^4 \overset{i_*}{\mapsto} & 0 & & \\
& c_2^1 \overset{i_*}{\mapsto} & c_2^2 \overset{i_*}{\mapsto} & 0 & & & & \\
& & & c_2^3 \overset{i_*}{\mapsto} & c_2^4 \overset{i_*}{\mapsto} & c_1^5 \overset{i_*}{\mapsto} & c_1^6 \overset{i_*}{\mapsto} & \cdots \\
& & & & c_3^4 \overset{i_*}{\mapsto} & c_2^5 \overset{i_*}{\mapsto} & 0 & \\
& & & & & c_3^5 \overset{i_*}{\mapsto} & 0 & \\
& & & & & & c_2^6 \overset{i_*}{\mapsto} & \cdots
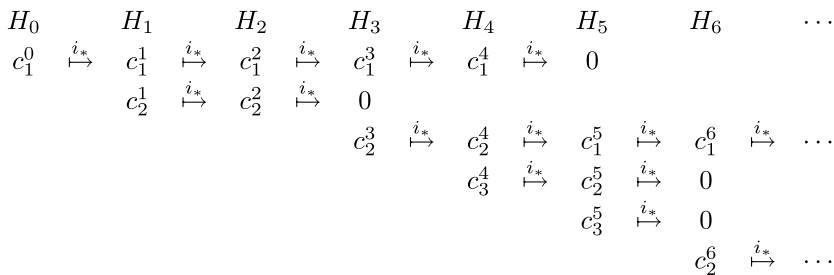\end{array}
$$

FIG 6.1. *Bases across the filtration.*

For any basis element c_k^j which is not an image of a previous basis element c_k^{j-1}, either there is a minimal number $p \geq 1$ such that $i_*^p(c_k^j) = 0$ or $i_*^p(c_k^j) \neq 0$ for any $p \geq 1$. Each such element can be matched to the interval $(j, j+p)$, or, respectively, (j, ∞). These intervals give rise to the graphical presentation in the form of a barcode. The horizontal axis of the barcode scheme represents time - i.e. the index of the complex - and each bar spanning the interval from j to i corresponds to an interval (j, i) linked to a basis element as above.

Thus, for the bases of Figure 6.1, assuming that $K_j = K_6$ for all $j \geq 6$, the barcode is as in Figure 6.2.

Note that the Betti numbers of each of the complexes in the filtration can be easily derived from its barcode: $\beta_n(K_j)$ is the number of bars which intersect a vertical line at time j, excluding those ending at time j.

Recall that we considered a single fixed dimension of the homology, n. The collections of bars of persistent homologies of all dimensions is called the barcode of the filtered complex (of finite type) $\mathcal{K} = \{K_j\}_{j \geq 0}$.

6.2. Random filtered complexes and entrance time fields

Random filtered complexes are the link between the 'deterministic' notion of persistent homology and the random setting.

Assuming the ubiquitous probability space $(\Omega, \mathcal{F}, P)$, a random filtered complex will be defined as a mapping from Ω to some space F of filtered complexes. However, allowing F to be too general makes it virtually impossible to define a meaningful σ-algebra on it, and so we shall restrict our discussion here to cases in which the elements of F are all sub-complexes of finite universal complex. Among other things, this implies that they are of finite type.

We begin with a trivial generalization of the definition of a filtered simplicial complex. We now allow the set of indices of a filtered complex to be any well-ordered infinite set in $\mathbb{R} \cup \{-\infty\}$, so that a filtered complex is now of the form
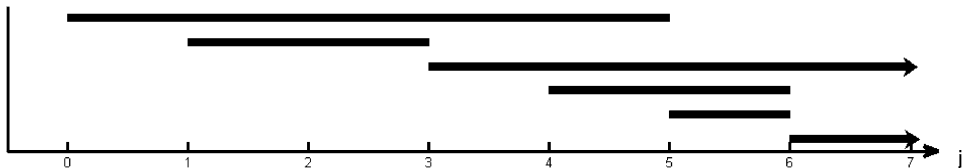

FIG 6.2. *Barcode representation of homology bases.*

$\mathcal{K} = \{K_\alpha\}_{\alpha \in A}$. The definitions which followed the definition of filtered complexes can be easily adjusted accordingly. The motivation for this is that, while in the deterministic case one deals with a fixed filtered complex, with a fixed set of indices, in the random case one needs to assign indices meaningful to a wider possible set of outcomes, and the natural numbers no longer suffice as an index set.

In addition, for the discussion of random filtered complexes, we think of finite type complexes as having only a finite number of indices by discarding the constant tail of complexes. Formally, we can think about it as restricting ourselves to discussing only complexes with tail defined in a canonical way: for all filtrations, $\{K_\alpha\}_{\alpha \in A}$, there exists $\alpha_0 \in A$ such that for all $\alpha_0 \leq \alpha \in A$, $K_{\alpha_0} = K_\alpha$ and such that $A \cap [\alpha_0, \infty) = \{\alpha_0 + i\}_{i=0}^{\infty}$. To save some tedious notation we simply work with a finite portion of the complex.

Next, some notation. For a given simplex σ in a filtered complex, we define its entrance time

$$\mathrm{ent} \;\equiv\; \mathrm{ent}(\sigma) \;\overset{\Delta}{=}\; \min\{\alpha \,\sigma \in K_\alpha\},$$

if the minimum is finite and ∞ otherwise. For a simplicial complex K, let $\mathfrak{F}(K)$ denote all finite type filtrations, $\{K_\alpha\}_{\alpha \in A}$, of K satisfying the condition that, for any $\alpha \in A$, there exists a simplex $\sigma \in K$ with entrance time $\mathrm{ent}(\sigma) = \alpha$. This condition basically says that we consider filtrations with no 'spare' complexes, which, loosely speaking, contain no additional information.

Note that we then have the natural injective mapping

$$\pi \;\equiv\; \pi_K \, \mathfrak{F}(K) \longrightarrow \prod_{\sigma \in K} \overline{\mathbb{R}}_\sigma = \overline{\mathbb{R}}^{\mathrm{card}(K)}, \;\; \{K_\alpha\}_{\alpha \in A} \longmapsto \{\mathrm{ent}(\sigma)\}_{\sigma \in K}.$$

Using the mapping π, $\mathfrak{F}(K)$ can be endowed with the structure of a measurable space, determined by the rule: $F \subset \mathfrak{F}(K)$ is measurable if and only if $F = \pi^{-1}(B)$ for some Borel set B in $\prod_{\sigma \in K} \overline{\mathbb{R}}_\sigma$, with respect to the standard product topology on $\prod_{\sigma \in K} \overline{\mathbb{R}}_\sigma$. We denote this σ-algebra by $\mathfrak{B}(K)$.

Note that $\mathfrak{B}(K)$ is the Borel σ-algebra on $\mathfrak{F}(K)$, when endowing it with the topology defined by the rule: $F \subset \mathfrak{F}(K)$ is open if and only if $F = \pi^{-1}(G)$ for some open set G in $\prod_{\sigma \in K} \overline{\mathbb{R}}_\sigma$.

We are now finally ready to define the random filtration of a complex.

Definition 6.4. Let K be a fixed universal complex. A random filtration of K is a measurable function $\mathcal{K} : (\Omega, \mathcal{F}) \to (\mathfrak{F}(K), \mathfrak{B}(K))$.

The following lemma, in which $\overline{\mathbb{R}}$ denotes the two point compactification of $\mathbb{R}$, is now straightforward.

Lemma 6.5. *For a finite complex K, the mapping $\mathcal{K} : (\Omega, \mathcal{F}) \to (\mathfrak{F}(K), \mathfrak{B}(K))$ is measurable if and only if $\mathrm{ent}_{\mathcal{K}(\omega)}(\sigma) : (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathfrak{B}(\overline{\mathbb{R}}))$ is measurable for all $\sigma \in K$ (where $\mathfrak{B}(\overline{\mathbb{R}})$ is the Borel σ-algebra on $\overline{\mathbb{R}}$).*

Lemma 6.5 implies that $\pi \circ \mathcal{K}$ is a random field on K. The next result shows that under certain compatibility conditions on a field on K, the converse is also true.

Definition 6.6. Let $E = \{E_\sigma\}_{\sigma \in K}$ be a random field on a finite simplicial complex K. If $E_\sigma(\omega) \leq E_\tau(\omega)$ for any simplices $\sigma, \tau \in K$ for which $\sigma \subset \tau$, we say that E is an entrance time field on K.

Corollary 6.7. *If $E = \{E_\sigma\}_{\sigma \in K}$ is an entrance time field on a finite simplicial complex K, then $\pi_K^{-1}(E)$ is a random filtration of K. Moreover, π_K gives a 1-1 correspondence between random filtrations and entrance time fields on K.*

Note that even when a field E' on K does not satisfy the condition of Definition 6.6 we can define an entrance time field by $E = \max\{E'_\tau : \tau \subset \sigma\}$.

References

[1] ABRAMS, A. and GHRIST, R. (2002). Finding topology in a factory: Configuration spaces. *Amer. Math. Monthly* **109** 140–150.

[2] ADLER, R. J. and TAYLOR, J. E. (2007). *Random Fields and Geometry*. Springer.

[3] BARYSHNIKOV, Y. and GHRIST, R. (2009). Euler integration over definable functions. Available at arXiv:0909.4054v1.

[4] BARYSHNIKOV, Y. and GHRIST, R. (2009). Target enumeration via Euler characteristic integrals. *SIAM J. Appl. Math.* **70** 825–844.

[5] BARYSHNIKOV, Y. and WEINBERGER, S. 2010. In preparation.

[6] BOBROWSKI, O. and BORMAN, M. S. (2010). Euler integration of Gaussian random fields and persistent homology. Available at arXiv:1003.5175.

[7] BOBROWSKI, O. and BORMAN, M. S. (2010). Expected number of critical points for the distance function and the topology of random simplicial complexes. In preparation.

[8] BUBENIK, P., CARLSSON, G., KIM, P. T. and LUO, Z. (2009). Statistical topology via Morse theory, persistence and nonparametric estimation. Available at arXiv:0908.3668.

[9] BUBENIK, P. and KIM, P. T. (2007). A statistical approach to persistent homology. *Homology, Homotopy and Applications* **9** 337–362.

[10] CARLSSON, G. (2009). Topology and data. *Bull. Amer. Math. Soc. (N. S.)* **46** 255–308.

[11] CARLSSON, G. (2010). Topological methods in scientific computing, statistics and computer science. Available at comptop.stanford.edu/.

[12] CARLSSON, G. and DE SILVA, V. Plex: MATLAB software for computing persistent homology of finite simplicial complexes. Available at comptop.stanford.edu/programs/plex.

[13] CARLSSON, G. and ZOMORODIAN, A. (2009). The theory of multidimensional persistence. *Discrete Comput. Geom.* **42** 71–93.

[14] CHAZAL, F., COHEN-STEINER, D. and MERIGOT, Q. (2009). Geometric inference for measures based on distance functions. Available at http://hal.inria.fr/inria-00383685/.

[15] CHUNG, M. K., BUBENIK, P. and KIM, P. T. (2009). Persistence diagrams of cortical surface data. In *Information Processing in Medical Imaging. Lecture Notes in Computer Science* **5636** 386–397. Springer, Berlin.

[16] DARPA (2010). Topological data analysis. Available at www.darpa.mil/dso/thrusts/math/funmath/tda/index.htm.

[17] de Silva, V. and GHRIST, R. (2007). Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.* **7** 339–358.

[18] DE SILVA, V. and GHRIST, R. (2007). Homological sensor networks. *Notices Amer. Math. Soc.* **54** 10–17.

[19] DIACONIS, P. From characterization to algorithm. MSRI lecture. Available at http://198.129.64.244/13156//13156-13156-Quicktime.mov.

[20] EDELSBRUNNER, H. and HARER, J. (2008). Persistent homology—a survey. In *Surveys on discrete and computational geometry, Contemp. Math.* **453** 257–282. Amer. Math. Soc., Providence, RI.

[21] FLATTO, L. and NEWMAN, D. J. (1977). Random coverings. *Acta Mathematica* **138** 241–264.

[22] FRISTON, K. (2010). Home page. Available at http://www.fil.ion.ucl.ac.uk/ ∼karl/.

[23] GHRIST, R. (2008). Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N. S.)*, 45(1):61–75 (electronic).

[24] GHRIST, R. (2008). Topological methods in electrical and systems engineering. Available at www.math.uiuc.edu/∼ghrist/notes/appltop/.

[25] GHRIST, R. (2009). Sensor Topology for Minimal Planning. Available at www. darpa.mil/dso/thrusts/math/funmath/stomp/index.htm.

[26] HATCHER, A. (2002). *Algebraic Topology*. Cambridge Univ. Press.

[27] HUGHES, B. and RANICKI, A. (1996). *Ends of Complexes. Cambridge Tracts in Mathematics* **123**. Cambridge Univ. Press, Cambridge.

[28] KAHLE, M. (2009). Random geometric complexes. Arxiv preprint. Available at arXiv:0910.1649.

[29] MECKE, K. R. and WAGNER, H. (1991). Euler characteristic and related measures for random geometric sets. *J. Statist. Phys.* **64** 843–850.

[30] NIYOGI, P., SMALE, S. and WEINBERGER, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** 419–441.

[31] NIYOGI, P., SMALE, S. and WEINBERGER, S. (2009). A topological view of unsupervised learning. Available at http://www.math.uchicago.edu/∼shmuel/ noise.pdf.

[32] OKUN, B. L. (1990). Euler characteristic in percolation theory. *J. Statist. Phys.* **59** 523–527.

[33] PENROSE, M. (2003). *Random Geometric Graphs*. Oxford Univ. Press.

[34] SUBAG, E. (2010). Algebraic topology and the morphology of fMRI images. I: Computational homology of random fields. Technion undergraduate project.

[35] TAYLOR, J. E. (2001). Euler characteristics for Gaussian fields on manifolds. PhD thesis, McGill Univ.

[36] TAYLOR, J. E. and ADLER, R. J. (2009). Poincaré's limit, kinematic formulae and Gaussian processes. *Ann. Probab.* **37** 1459–1482.

[37] VICK, J. W. (1973). *Homology Theory*. Academic Press, New York.

[38] VIRO, O. Y. (1988). Some integral calculus based on Euler characteristic. *Lecture Notes in Math.* **1346** 127–138.

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