Persistent homology and the topology of motor cortical activity

Julian Salazar and Emma West
jsalazar01@college.harvard.edu · emma_west@g.harvard.edu

Abstract. We introduce the underlying principles of topological data analysis via persistent homology. We then preview how these can be applied to analyze the activity-based topologies of various neuronal classes in the motor cortex, along with their evolution as induced by engagement in a physical task.

Contents

1. Introduction

1.1. Topological data analysis

Topology is the branch of mathematics which studies the qualitative geometry of a space. This includes information about connectivity and higher-dimensional structures such as holes and voids, attributes that are preserved under homeomorphisms like continuous deformation. In recent years, the field of **topological data analysis (TDA)** has emerged as a way to employ topological tools such as homology to point clouds of discrete data. This technique is particularly useful in recovering qualitative information about a data set that might not be recognized using conventional analytic methods, as these methods often make assumptions of distance-dependence and linear separability. TDA generally returns a summary of

the overall data structure that is insensitive to coordinates and metric, via examination of the relationships between functorial geometric constructions that arise from the data [Car09]. For example, TDA shines when describing data that has non-trivial underlying topology (Figure 1). One might imagine that the data is sampled from a manifold embedded in $\mathbb{R}^N$, up to the presence of some noise.

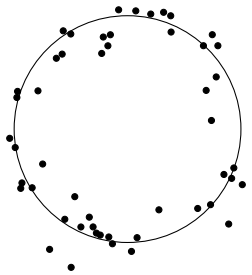


FIGURE 1. A data set in $\mathbb{R}^2$ with the underlying topology of a circle. TDA would aim to detect a single connected component with a 1-dimensional hole, independent of continuous deformation.

One of the primary methods of TDA is **persistent homology** (Section 2.3), which seeks to understand the topology underlying a point cloud data set at various resolutions. We can apply this method by building an intermediate space, the Vietoris-Rips complex (Definition 2.23), which depends on a parameter $\epsilon > 0$ and on the notion of "distance" between any two data points. This is given by the choice of metric on the space, such as the one given by embedding into an ambient space like $\mathbb{R}^N$ and using Euclidean distance.

The Rips complex is built by connecting all pairs of points x and y such that $d(x,y) < \epsilon$ via an edge, all triplets x, y, z such that

$$d(x,y) < \epsilon, \quad d(x,z) < \epsilon, \quad d(y,z) < \epsilon$$

via filled triangles, all quadruplets with pairwise distances less than ϵ via filled tetrahedra, and so forth. Changing the value of ϵ gives different Rips complexes for which invariants known as Betti numbers (Definition 2.20) can be computed. Rather than restricting to a single spatial scale and risking the loss of valuable information, persistent homology refers to the properties which persist as the topology evolves along different spatial scales (parameterized by ϵ). If the underlying manifold of data has multidimensional "holes", we expect to detect them over some interval of spatial scale, that is, for some interval of ϵ within 0 to ∞. The holes that persist through relatively long ranges of scales are thought to represent important features of the data [Wol16].

The robust nature of TDA is appealing for the analysis of biological data due to the intrinsic variability and stochastic nature of biological processes, and the fluid definitions of biological similarity. In 2011, a group of scientists and mathematicians demonstrated how TDA can be useful in transcriptional data analysis by using this method to classify a new kind of breast cancer based on a distinct and statistically significant molecular signature [NLC11].

Another remarkable application of TDA was carried out in 1979 by Miller and Reaven on data from 145 diabetic patients [RM79]. Analyzing both diabetics

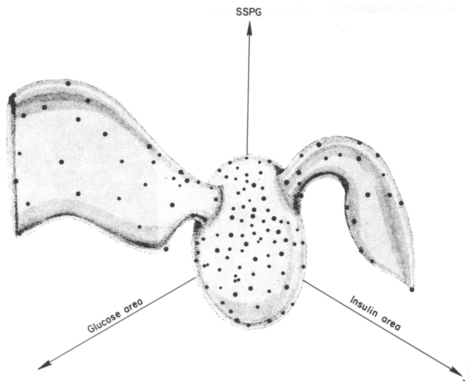


FIGURE 2. This figure, taken from [RM79] is an artist's depiction of the data in three dimensions. The central region represents non-diabetic patients, the left cluster Type I diabetes, and the right cluster Type II.

and non-diabetics, four metabolic variables pertinent to glucose tolerance and insulin response were recorded, as well as patient weight. This constructed 145 data points in $\mathbb{R}^5$ - a significantly large data set with no conventional or existing method of analysis. Using the Prim 9 program to project this data into 3-dimensional space (see Figure 2), it became apparent that the subjects with diabetes likely belonged to different populations and ultimately, this elucidated the classification of diabetes into Type I (left region) and Type II (right region).

1.2. Motor cortical activity

In this study, we have chosen to apply topological data analysis to understand motor cortical activity. Thanks to the work of Kazutaka Takahashi and the Hastopoulos lab at the University of Chicago, aggregate signals in the motor cortex have been understood as propagating waves across the cortical surface [TKC+15], but far less is understood about the spatial arrangement and coordination of individual neurons and about how the signaling of individual neurons contributes to the aggregate signals. This group has shown that a large network of neurons display coordinated activation at movement onset, based on statistical classification of these neurons, and they claim that this suggests the necessity for a unique pattern of spatial activation for movement onset to occur [TKC+16]. The study also suggested that during movement onset, there appeared to be an increase in the number of statistically significant connections between the subclass of neurons with firing activity correlated to the aggregate signal, but no significant increase in the connectivity between the other class of neurons.

However, their analysis thus far is relatively constrained to observations like how the flow of information via the connections of individual neurons is directed in a similar way to the aggregate wave direction. While this is interesting, our hope is that finer spatial properties can be teased out as well, given the highly non-trivial structure of neural connectivity.

We turn to TDA since, as stated before, it represents a fundamentally different approach that can qualify intrinsic properties of an underlying space that are invariant under choice of metric and/or continuous deformation. This is desirable in the context of neurons and their activity (intuitively, we think of differences in functional and physical structure between brains as the types of "noisy" deformation we wish to ignore). By constructing a mathematical space from the activity of motor cortical neurons (Section 4.2) we can look at the same neuronal data in a new light, with the hope of understanding the topological signatures or differences that arise from a set of neurons and how that topology changes in time relative to movement onset.

Our precedent is a 2010 study where mathematicians applied persistent homology to data derived from an electrode array in the primary visual cortex of macaque monkeys [SMI+08]. They constructed data sets from the simultaneous activity of a number of neurons to get a qualitative understanding of their population's interaction during visual stimulation. In particular, they sought to "offer an estimate of the underlying topological structure of V1 activity." Their method of analysis, as well as our own, will be discussed in Section 4.

2. Mathematical background

To apply the technique of persistent homology, one first needs to convert point cloud data into a filtration of spaces that can be meaningfully analyzed using algebraic topology. The standard technique involves turning point-cloud data into simplicial complexes known as Vietoris-Rips complexes, and subsequently computing their respective simplicial homologies. Here, we develop the necessary mathematical objects and describe the relevant results needed for our later topological analysis. This section synthesizes information from [Hat02] (a standard text in algebraic topology), along with a guide to computing persistent homology released by two pioneers in the field [ZC05].

2.1. Simplicial complexes

Definition 2.1. Let $v_0,\dots,v_n$ be points in $\mathbb{R}^m$. Then:

- A point $x = \sum_{i=0}^{n} \lambda_i v_i$ with $\lambda_i \in \mathbb{R}$ is an **affine combination** of the v_i if $\sum_{i=0}^{n} \lambda_i = 1$. It is a **convex combination** if all $\lambda_i \geq 0$.
- A set of points is **affinely independent** when any two affine combinations are the same if and only if their corresponding coefficients in $\mathbb{R}$ are identical. This is equivalent to the difference vectors

$$v_1 - v_0,\ \dots,\ v_n - v_0$$

being linearly independent.
- The **convex hull** of $v_0,\dots,v_n$ is the set of their convex combinations:

$$x = \sum \lambda_i v_i \text{ where } \lambda_i \geq 0 \text{ for all } i \text{ and } \sum \lambda_i = 1$$

All together, these let us define the central object of (geometric) simplicial homology:

Definition 2.2. An n-**simplex** is the convex hull of $n+1$ affinely independent points. The points v_i are the vertices of the simplex and the simplex is denoted by $[v_0,\dots,v_n]$. We say n is the **dimension** of the simplex.

Example 2.3. The **standard n-simplex** Δ^n is that whose vertices are the unit vectors along the coordinate axes:

$$\Delta^n = \left\{ (t_0, \ldots, t_n) \in \mathbb{R}^{n+1} \;\middle|\; \sum_{i=0}^{n} t_i = 1 \text{ and } t_i \geq 0 \text{ for all } i \right\}$$

For the purpose of developing simplicial chain complexes, we will require the following notion:

Definition 2.4. An **ordered n-simplex** is an n-simplex with a total order on its vertices. We write $[v_0, \ldots, v_n]$ if the ordering is

$$v_0 < v_1 < \ldots < v_{n-1} < v_n.$$

The ordering of the vertices will impose **orientations** on each edge, orienting $[v_i, v_j]$ in the direction of increasing subscripts.

Note that the linear independence of the n vectors $v_1 - v_0$, …, $v_n - v_0$ implies that an n-simplex is indeed homeomorphic to a closed n-ball. Intuitively, an n-simplex is the filled n-dimensional analog of a triangle. The first few dimensions are visualizable: a 0-simplex is simply a point, a 1-simplex is an edge, a 2-simplex is a filled triangle, and a 3-simplex is a filled tetrahedron (see Figure 3 for oriented examples).

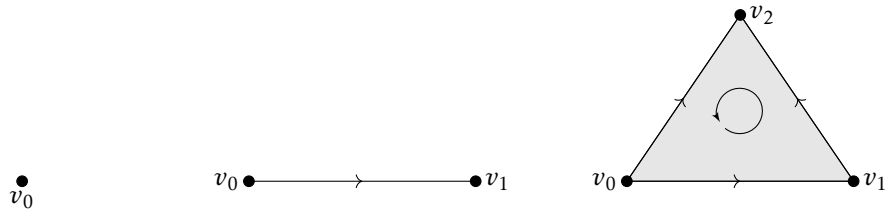


FIGURE 3. Examples of oriented 0, 1, and 2-simplices, respectively

Definition 2.5. Eliminating one of the $n+1$ vertices of an n-simplex results in n vertices that span an $(n-1)$-simplex, called a **face**.

Note that if the larger simplex is ordered, the vertices of a face receive an induced order which also make it an ordered simplex. We see this occur for the 2-simplex in Figure 3: $[v_0, v_1, v_2]$ gives faces $[v_0, v_1]$, $[v_0, v_2]$, and $[v_1, v_2]$.

Definition 2.6. The **boundary** of an n-simplex σ, denoted $\partial\sigma$, is the union of all of σ's faces. The remaining portion of the simplex, which is the interior of σ is called the **open simplex**. It is given by the set difference $\sigma - \partial\sigma$.

We form simplicial complexes by identifying simplices together along their faces, as formalized in the following definition:

Definition 2.7. A finite set of simplices K is a **simplicial complex** if the following properties hold:

a) For all simplices $\sigma \in K$, if τ is a face of σ, then $\tau \in K$ as well.
b) $\sigma, \tau \in K \implies \sigma \cap \tau$ is empty or a simplex itself.

The **dimension** of K is the maximum dimension across all its simplices.

Given a simplicial complex K, we can throw away the geometric interpretation of the simplices by retaining only the set of vertices of each simplex, along with the implicit combinatorial structure given by the relationship between vertex sets. This gives an example of an abstract simplicial complex A, which we call the **vertex scheme** of K. Formally:

Definition 2.8. An **abstract simplicial complex** is a finite collection A of finite subsets of a vertex set V, such that each element of V is a simplex and $\alpha \in A$ and $\beta \subseteq \alpha$ implies that $\beta \in A$.

We then call sets in A our **simplices**, where the dimension of a simplex is $\dim(\alpha) = \text{card}(\alpha) - 1$. The dimension of the complex is still the maximum dimension over its simplices, and the vertex set of A is the union of all 0-simplices in A.

Abstract simplicial complexes can be thought of as an abstract version of a geometric simplicial complex. We say that K is a **geometric realization** of A if its vertex scheme is isomorphic to A. The other direction of generating a geometric realization from an abstract simplicial complex is slightly more difficult, but it can be done in a sufficiently high-dimensional Euclidean space by the following theorem, with proof adapted from [Ede06]:

Theorem 2.9 (Geometric realization theorem)**.** *An abstract simplicial complex A of dimension d has a geometric realization in $\mathbb{R}^{2d+1}$.*

Proof. Let $f : \text{Vert } A \to \mathbb{R}^{2d+1}$ be an injection whose image is a set of points in **general position** (i.e., any $2d+2$ or fewer of the points are affinely independent). Take α, α_0 to be simplices in A with $k = \dim(\alpha)$ and $k_0 = \dim(\alpha_0)$. The union of the two simplices has size:

$$\text{card}(\alpha \cup \alpha_0) = \text{card}(\alpha) + \text{card}(\alpha_0) - \text{card}(\alpha \cap \alpha_0) \leq k + k_0 + 2 \leq 2d + 2$$

Therefore, the points in $\alpha \cup \alpha_0$ are affinely independent, implying that every convex combination x of points in this space is unique. Thus, x is in $\sigma = \text{conv}(f(\alpha))$ and x is in $\sigma_0 = \text{conv}(f(\alpha_0))$ if and only if x is a convex combination of $\alpha \cap \alpha_0$. But this indicates that $\sigma \cap \sigma_0 = \emptyset$ or $\text{conv}(f(\alpha \cap \alpha_0))$, completing the proof. □

2.2. Homology groups

We now work towards defining the simplicial homology groups of a simplicial complex K. Let $C_p(K)$ denote the free abelian group over the oriented p-simplices $\{\sigma_j^p\}$ of K.

Definition 2.10. An element $X \in C_p(K)$ is called a p-**chain**. It is a formal sum of p-simplices in K with coefficients in $\mathbb{Z}$. That is, X is of the form

$$X = \alpha_1 \sigma_1^p + \ldots + \alpha_k \sigma_k^p,$$

where $\{\alpha_j \mid j = 1, \ldots, k\} \in \mathbb{Z}$. It follows that $C_p(K)$ is an abelian group under addition.

The boundary of an n-simplex $[v_0, \ldots, v_n]$ is constituted by the $(n-1)$-simplices $[v_0, \ldots, \hat{v_i}, \ldots, v_n]$, i.e., the simplices resulting from removing the i-th vertex for each i. In the language of chains, the boundary is the $(n-1)$-chain given by an "oriented" sum of the faces $[v_0, \ldots, \hat{v_i}, \ldots, v_n]$. This can be formalized as follows:

Definition 2.11. Given a simplicial complex K, we define a **boundary homomorphism** ∂_p. We specify this map for all values of $C_p(K)$ by specifying its value on basis elements. Let $p > 0$ and $\sigma = [v_0, \ldots, v_p]$ be an oriented p-simplex. Then:

$$\partial_p : C_p(K) \to C_{p-1}(K), \quad \partial_p \sigma = \sum_{i=1}^{p} (-1)^i [v_0, \ldots \hat{v_i} \ldots, v_p]$$

where $\hat{v_i}$ means the vertex has been omitted. The image of a p-chain under ∂_p is the p-chain's **boundary**.

Example 2.12. Consider once more the example of oriented simplices in Figure 3. Using the definition of the boundary homomorphism above, we compute the boundaries for the 1 and 2-simplices:

$$\partial[v_0, v_1] = [v_1] - [v_0]$$
$$\partial[v_0, v_1, v_2] = [v_1, v_2] - [v_0, v_2] + [v_0, v_1],$$

where we can think of $-[v_0, v_2]$ as preserving the orientation of the edge from v_2 to v_0.

Lemma 2.13. *For any p-simplex σ, $(\partial_{p-1} \circ \partial_p)\sigma = 0$.*

Proof. Take $\sigma = [v_0, \ldots, v_p]$. Then:

$$\begin{aligned}
(\partial_{p-1} \circ \partial_p)\sigma &= \partial_{p-1}\Big(\sum_{i=0}^{p} (-1)^i [v_0, \ldots, \hat{v_i} \ldots, v_p] \Big) \\
&= \sum_{j=0}^{i-1} (-1)^i [v_0, \ldots, \hat{v_j}, \ldots, \hat{v_i} \ldots, v_p] + \sum_{j=i}^{p} (-1)^i [v_0, \ldots, \hat{v_i}, \ldots, \hat{v_j} \ldots, v_p] \\
&= \sum_{i=0}^{p} \sum_{j=0}^{i-1} (-1)^{i+j} [v_0, \ldots, \hat{v_j}, \ldots, \hat{v_i} \ldots, v_p] + \sum_{i=0}^{p} \sum_{j=i+1}^{p} (-1)^{i+j-1} [v_0, \ldots, \hat{v_i}, \ldots, \hat{v_j} \ldots, v_p] \\
&= \sum_{j<i} (-1)^{i+j} [v_0, \ldots, \hat{v_j}, \ldots, \hat{v_i} \ldots, v_p] + \sum_{j>i} (-1)^{i+j-1} [v_0, \ldots, \hat{v_i}, \ldots, \hat{v_j} \ldots, v_p] \\
&= 0.
\end{aligned}$$

This follows because we can switch the roles of i and j in the last sum. When we do so, the second sum becomes the negative of the first, completing our proof. □

For all p, recall that $C_p(K)$ forms an abelian group. We now have a sequence of homomorphisms of abelian groups with the property that $\partial_{p-1} \circ \partial_p = 0$. Such a sequence is formally called a **chain complex**.

$$C_p(K) \xrightarrow{\partial_p} C_{p-1}(K) \xrightarrow{\partial_{p-1}} C_{p-2}(K) \longrightarrow \ldots \longrightarrow C_1(K) \xrightarrow{\partial_1} C_0(K) \xrightarrow{\partial_0} 0$$

Definition 2.14. A p-chain is a **cycle** if its boundary is zero. Let

$$Z_p = \{p\text{-cycles of } K \text{ over } Z\}$$

Then $Z_p(K)$ is a subgroup of $C_p(K)$ and by definition:

$$Z_p(K) = \mathrm{Ker}(\partial_p)$$

An interesting way to understand cycles is to note that the boundary of any simplex is, in fact, a cycle. This follows directly from our lemma $\partial_{p-1} \circ \partial_p = 0$, i.e., that the formal sum of the boundaries of the boundary of any simplex is always zero.

Definition 2.15. A p-chain is a **boundary** if it can be written as the boundary of an element in $C_{p+1}(K)$. Let $B_p(K)$ denote this set of boundaries. Therefore:

$$B_p(K) = \mathrm{Im}(\partial_{p+1})$$

From this, since $\partial_{p-1} \circ \partial_p = 0$, it must be true that $B_p(K) \subset Z_p(K)$. Finally, we will use this to define the homology groups of K.

Definition 2.16. The p-**th homology group** of a simplicial complex K is given by:

$$H_p(K) = Z_p(K)/B_p(K) = \mathrm{Ker}(\partial_p)/\mathrm{Im}(\partial_{p+1})$$

This quotient is valid since we are looking at subgroups of the abelian group $C_p(K)$, which ensures that $B_p(K)$ is indeed a normal subgroup of $Z_p(K)$. The elements of $H_p(K)$ are equivalence classes of p-cycles that do not bound any $p+1$ chain. In this way, the homology group characterizes p-dimensional holes in the original complex. Two p-cycles are in the same equivalence class if $z_1, z_2 \in Z_p(K)$ such that $z_1 - z_2 \in B_p(K)$. If this holds, then z_1 and z_2 are considered **homologous**.

For a finite simplicial complex, the homology groups are finitely generated abelian groups, which means we can understand their structure more generally according to the following theorem:

Theorem 2.17. *If G is a finitely generated abelian group then it is isomorphic to a direct sum of the form*

$$G \cong \mathbb{Z}^\beta \oplus \mathbb{Z}/t_1 \oplus \cdots \oplus \mathbb{Z}/t_m.$$

for non-negative β and positive $t_1, \ldots, t_m$.

In general, one can compute homology with coefficients in a group G. This involves starting with the tensor product $C_p(K) \otimes_{\mathbb{Z}} G$, which can viewed as formal sums

$$X = \alpha_1 \sigma_1^p + \ldots + \alpha_k \sigma_k^p,$$

as before, but with $\alpha_1, \ldots, \alpha_k \in G$.

Definition 2.18. The p-**th homology group with coefficients in** G of a simplicial complex K, written $H_p(K; G)$, comes from instead taking

$$C_p(K) \otimes G \xrightarrow{\partial_p \otimes \mathrm{id}_G} C_{p-1}(K) \otimes G \longrightarrow \ldots \longrightarrow C_1(K) \otimes G \xrightarrow{\partial_1 \otimes \mathrm{id}_G} C_0(K) \otimes G \xrightarrow{\partial_0 \otimes \mathrm{id}_G} 0,$$

defining $Z_p(K; G)$ and $B_p(K; G)$ with respect to these $\partial_p \otimes \mathrm{id}_G$'s, and then taking

$$H_p(K; G) = Z_p(K; G)/B_p(K; G) = \mathrm{Ker}(\partial_p \otimes \mathrm{id}_G)/\mathrm{Im}(\partial_{p+1} \otimes \mathrm{id}_G).$$

Remark. In this notation, $H_p(K; \mathbb{Z}) = H_p(K)$.

Example 2.19. Most TDA practitioners take $G = \mathbb{Z}/2\mathbb{Z}$, which we also view as the additive group of the field $\mathbb{F}_2$. This has the effect of discarding orientation, since now $-\sigma = \sigma$. This weakens our ability to distinguish between topological spaces,

but greatly simplifies computation. By the *universal coefficient theorem of homology*, we get a split exact sequence showing the existence of an isomorphism:

$$H_p(K;\mathbb{Z}/2\mathbb{Z}) \cong (H_p(K)\otimes\mathbb{Z}/2\mathbb{Z})\oplus \mathrm{Tor}(H_{p-1}(K),\mathbb{Z}/2\mathbb{Z}).$$

(See [Hat02, §3A] for details on the theorem and the Tor functor.) In practical terms, if via Theorem 2.17 we write our (finitely-generated) homology groups as

$$H_p(K) \cong \mathbb{Z}^\beta \oplus \mathbb{Z}/t_1 \oplus \cdots \oplus \mathbb{Z}/t_m$$
$$H_{p-1}(K) \cong \mathbb{Z}^\gamma \oplus \mathbb{Z}/u_1 \oplus \cdots \oplus \mathbb{Z}/u_\ell,$$

then our isomorphism shows that

$$H_p(K;\mathbb{Z}/2\mathbb{Z}) \cong (\mathbb{Z}/2\mathbb{Z})^{\beta+t} \oplus (\mathbb{Z}/2\mathbb{Z})^u \cong (\mathbb{Z}/2\mathbb{Z})^{\beta+t+u},$$

where t, u are the number of t_i's, u_j's that are divisible by 2, respectively. This is a result on the rank, and so the actual choice of isomorphism is unimportant; when doing computations, one would compute $H_p(K;\mathbb{Z}/2\mathbb{Z})$ outright instead of computing $H_p(K)$ and passing to $H_p(K;\mathbb{Z}/2\mathbb{Z})$ afterwards.

Most importantly, $H_p(K;\mathbb{Z}/2\mathbb{Z})$ is a $\mathbb{Z}/2\mathbb{Z}$-vector space. This leads to a sequence of numbers with useful geometric interpretation:

Definition 2.20. The dimension of $H_p(K;\mathbb{Z}/2\mathbb{Z})$ is the **p-th Betti number of K (with coefficients in $\mathbb{Z}/2\mathbb{Z}$)**, written $\beta_p = \beta + t + u$. For any space, these Betti numbers over $\mathbb{Z}/2\mathbb{Z}$ form a sequence of non-negative integers $(\beta_0, \beta_1, \beta_2, \ldots)$ which we call the **Betti signature**.

Betti signatures are **topological invariants**, as spaces which are **homeomorphic** (i.e., there exists a bicontinuous map transforming one to the other) have the same Betti signatures. For example, if K is a simplicial complex, then the complex K' where some vertex v was shifted by a small ϵ will still have the same homology groups and Betti signature, despite being a different space.

Remark. How does one interpret the Betti signature geometrically? The mantra is that $H_p(K)$ consists of "cycles mod boundaries" (where working in $H_p(K;\mathbb{Z}/2\mathbb{Z})$ causes some of these cycle classes to vanish). Thus, the dimension β_p of $H_p(K;\mathbb{Z}/2\mathbb{Z})$ roughly expresses the maximal number of *independent* cycles that are not boundaries. In the 1-dimensional setting, this corresponds to the number of holes (see Figure 4).

Example 2.21. If two vertices (0-simplices) are connected by a sequence of 1-simplices, then they are the boundary of the corresponding 1-chain. Hence β_0 counts **the number of connected components**.

Example 2.22. A higher-dimensional example is the torus, which has Betti signature $(1, 2, 1, 0, \ldots)$. One can triangulate the torus to get a simplicial complex comprised by 2-simplices. Then:

- The 2-chain summing over all the 2-simplices is a cycle (each triangle edge σ, is shared by exactly 2 triangles, and thus upon taking boundaries gives 0). This 2-chain is a generator that gives $\beta_2 = 1$, as it encloses the "void" that is the torus' interior.
- One can find a basis of two 1-chains to give $\beta_1 = 2$. One is any sum of 1-simplices corresponding to a meridian of the torus. The other is any sum corresponding to the equator of the torus.

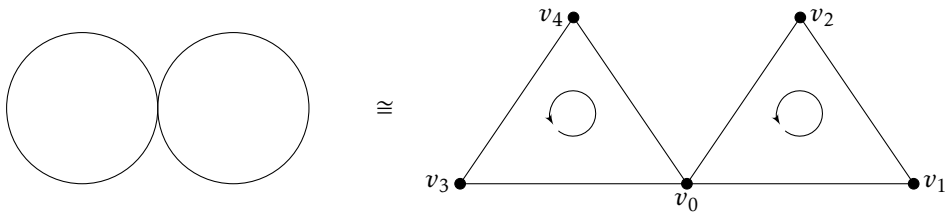


FIGURE 4. The wedge of two circles is homeomorphic to a simplicial complex with Betti signature $(1, 2, 0, \dots)$. For example, $\beta_1 = 2$ expresses that $H_p(K; \mathbb{Z}/2\mathbb{Z})$ has the basis $[v_0, v_1] + [v_1, v_2] + [v_0, v_2]$ and $[v_0, v_3] + [v_3, v_4] + [v_0, v_4]$, which are cycles without being boundaries of some 2-chain.

- The torus (and its triangulation) is a connected space, so by Example 2.21 we have $\beta_0 = 1$.

2.3. Persistent homology

In the previous section, we defined simplicial homology and showed how Betti numbers are a geometrically-descriptive topological property of a simplicial complex. In this section, we will introduce the application of simplicial homology to topological data analysis of point cloud data. The proposed method, called persistent homology, builds a simplicial complex from the point cloud data and analyzes the homology of the complex. One hopes that the constructed simplicial complex approximates the hypothetical underlying space from which our points are sampled from.

We study a specific complex construction that is induced by three pieces of data:

- A discrete set of points $\{x_\alpha\}$ in a topological space X (for example, $X = \mathbb{R}^n$).
- A distance function (metric) $d : X \times X \to \mathbb{R}_{\geq 0}$
- A choice of $\epsilon > 0$.

One creates an abstract simplicial complex as follows (description courtesy of [Sal14]):

Definition 2.23. Given the discrete points $\{x_\alpha\} \subseteq X$ with metric d and choice of $\epsilon > 0$, we create the **Vietoris-Rips complex** $\mathcal{VR}_\epsilon$, an abstract simplicial complex, by taking $\{x_\alpha\}$ to be our vertex set, and creating abstract k-simplices $[x_0, x_1, \dots, x_k]$ for every $(k+1)$-subset in $\{x_\alpha\}$ such that:

$$d(x_i, x_j) < \epsilon \text{ for all } 0 \leq i, j \leq k.$$

(This definition ensures that subsets of our abstract k-simplices are also simplices.)

By the geometric realization theorem (Theorem 2.9), the abstract simplicial complex $\mathcal{VR}_\epsilon$ can be embedded in some $\mathbb{R}^N$, allowing for practical visualization (even if our original space X is not obviously a subset of Euclidean space, such as a space of functions). Furthermore, by the correspondence of finite abstract and geometric simplicial complexes, if we were taking $X = \mathbb{R}^N$ already, then we could also perform the Vietoris-Rips construction geometrically (by taking geometric k-simplices in the space instead). We do this in the following example:

Example 2.24. Consider Figure 5, which illustrates three points in $\mathbb{R}^2$. The first row depicts each point enclosed by a ball of radius $\epsilon/2$, for small ϵ (left) and large ϵ (right). Beneath these are the corresponding Vietoris-Rips complexes, realized geometrically (the points are taken to be 0-simplices).

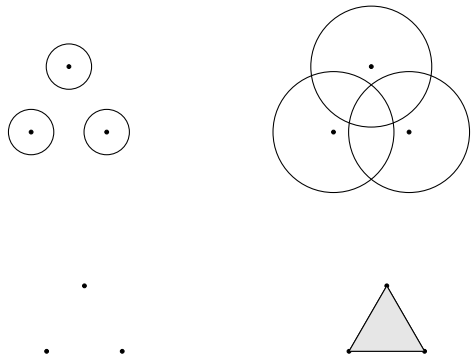


FIGURE 5. Simplicial complex and Vietoris-Rips complexes for varying ϵ.

On the left, ϵ is too small for any points to be within distance ϵ pairwise, resulting in no higher-dimensional simplices. However on the right, since the circles intersect pairwise, then the points are less than ϵ apart, and we add corresponding 1- and 2-simplices.

Remark. Note that if the points had been arranged in a square, then for some intermediate value of ϵ, the 1-simplices would be created but not the 2-simplices.

Finally, by varying the choice of ϵ we induce a **filtration**, a nested sequence of subcomplexes [Ede06]. In particular, if $\epsilon < \epsilon'$, then $\mathcal{VR}_\epsilon$ is a subcomplex of $\mathcal{VR}_{\epsilon'}$ and for any increasing sequence $\{\epsilon_i\}_0^n$, there exist natural inclusion maps:

$$\mathcal{VR}_{\epsilon_0} \hookrightarrow \mathcal{VR}_{\epsilon_1} \hookrightarrow \ldots \hookrightarrow \mathcal{VR}_{\epsilon_n}$$

The inclusion maps $\mathcal{VR}_{\epsilon_i} \to \mathcal{VR}_{\epsilon_j}$ for $i \leq j$ are continuous simplicial maps, i.e., they send a k-simplex of $\mathcal{VR}_{\epsilon_i}$ to a k'-simplex of $\mathcal{VR}_{\epsilon_j}$ where $k' = k$. As it sends cycles and boundaries to cycles and boundaries respectively, this defines corresponding homomorphisms on the p-th homology groups with coefficients in G (in general, this follows since homology is a *functor*):

$$H_p(VR_{\epsilon_0}; G) \to H_p(VR_{\epsilon_1}; G) \to \cdots \to H_p(VR_{\epsilon_n}; G)$$

This sequence of maps are called the **induced homomorphisms**, and we label them as:

$$f_p^{i,j} : H_p(VR_{\epsilon_i}; G) \to H_p(VR_{\epsilon_j}; G)$$

where $0 \leq i < j \leq n$ for any dimension p.

From these maps, we define the analogs for homology groups and Betti numbers for persistence:

Definition 2.25. The p-**th persistent homology groups** (with coefficients in G) are the images of $f_p^{i,j}$ for all $0 \leq i < j \leq n$:

$$\begin{aligned} H_p^{i,j} &= \mathrm{Im}\left(f_p^{i,j}\right) \\ &= Z_p(\mathcal{VR}_{\epsilon_i}; G)/(B_p(\mathcal{VR}_{\epsilon_j}; G) \cap Z_p(\mathcal{VR}_{\epsilon_i}; G)) \end{aligned}$$

If we take G to be $\mathbb{Z}/2\mathbb{Z}$, then $H_p^{i,j}$ is again a $\mathbb{Z}/2\mathbb{Z}$-vector space. Then as before, we can take the p-**th persistent Betti numbers** as:

$$\beta_p^{i,j} = \mathrm{rank}(H_p^{i,j}).$$

Note that the map $f_p^{i,j}$ sends homology classes of $\mathcal{VR}_{\epsilon_i}$ to those of $\mathcal{VR}_{\epsilon_j}$ and in this transition, some classes are subsumed into other classes or die out. In persistent homology, we are interested in the classes which persist under the map and we look at $H_p^{i,j}$ to gather this information. In particular, the classes that exist in $\mathcal{VR}_{\epsilon_i}$ and survive to exist in $\mathcal{VR}_{\epsilon_j}$ are classes of cycles that do not become boundaries in $\mathcal{VR}_{\epsilon_j}$, and $\beta_p^{i,j}$ gives the dimension of this vector space of persisting cycle classes.

A standard method of visualizing persistent Betti numbers was developed by Edelsbrunner, Letscher, and Zomorodian [HE00], which produces a graph of intervals for each homology class in a given dimension. Intervals are plotted above an axis parameterizing ϵ, where a non-trivial cycle class that begins in $\mathcal{VR}_{\epsilon_i}$ and dies at $\mathcal{VR}_{\epsilon_j}$ corresponds to a visual bar depicting the interval $[\epsilon_i, \epsilon_j]$. A set of such intervals is called the **barcode** for that dimension. Classes that persist over large intervals of ϵ are considered topological signatures of the purported underlying space, whereas short-lived classes may be inherent noise (as the data represents discrete samples). Examples can be found at the end of this paper, in Figure 8.

Analyzing a dataset using persistent homology involves creating the Vietoris-Rips complex at a series of incremented spatial scales. In this way, persistent Betti numbers identify lasting features of the data, providing a improved understanding of a hypothesized underlying topological space, one not limited to the structure imputed at a single spatial scale.

Remark. Namely, one hopes that (one of) the long-lasting i, j-signatures of persistent Betti numbers $(\beta_0^{i,j}, \beta_1^{i,j}, \dots)$ (where long-lasting means $\epsilon_j - \epsilon_i$ is relatively large) coincides with the Betti signature of the underlying space.

2.4. Witness complexes

Sadly, however, there are pragmatic considerations that prevent us from actually computing the full Vietoris-Rips complex for most values of ϵ. The Vietoris-Rips complex can become computationally unreasonable for large data sets relatively quickly - for n points in $\mathbb{R}^m$, the complex can have as many as $\Omega(n^m)$ simplices [Dey13]. To work around this computational hurdle, Carlsson and de Silva proposed a method for building a streamlined simplicial complex from a data set that can be made significantly smaller than the Vietoris-Rips complex by strategically selecting only a subset of data points, called **landmark points** [DSC04]. Performing persistent homology on this witness complex is far more computationally efficient and can be proven a robust alternative. The following construction of the witness complex, originally proposed in [DSC04] has been adapted from [Hen15].

Definition 2.26. A **witness complex** $\mathcal{W}(Z,L)$ is a simplicial complex constructed from a point cloud data set Z and a subset of **landmark points** L as follows:
The simplex $\sigma = [\sigma_1,\ldots,\sigma_k] \subset L$ is contained in $\mathcal{W}(Z,L) \Leftrightarrow$ there is a witness point $x \in Z$ such that:

$$\forall \sigma_i \in \sigma, \forall y \in L \setminus \{\sigma_1,\ldots,\sigma_k\}, \text{ we have } d(x,\sigma_i) < d(x,y).$$

There are two standard ways to choose the landmark set, each with its own horrors:

- **Random selection:** Let t be the number of landmark points desired. Choose t points from Z at random.
- **Max-min selection:** Let t be the number of landmark points desired. Begin by selecting $p_1 \in Z$ at random. Using induction, suppose $p_1, p_2, \ldots, p_{i-1}$ have already been chosen. Take $p_i \in Z$ be a point that maximizes the function:

  $$x \mapsto \min\{d(x,p_1), d(x,p_2), \ldots, d(x,p_{i-1})\}$$

  Repeat this inductively until t points have been selected.

Random selection tends to recover a landmark set that is reflective of the inherent density of Z. The **max-min** method tends to choose landmark points that are well-distributed within the space, but is more susceptible to noise and tends to select outliers, which may result in an unrepresentative topology. Both methods construct computationally efficient witness complexes that retain the underlying topology of the original space up to some error, making them appealing options for high-dimensional data analysis. For more on the justification of the witness complex as a reliable representation of the data space, see [DSC04].

3. Data collection and classification

In this section, we describe the nature of the experimental data made available for our topological analysis. The experimental conditions and some of the subsequent data analyses are outlined in significant detail in the prior work of Takahashi et al. [TKC+15], which we summarize in this section. In addition, we perform further types of classifications to identify additional neuronal classes for later use. The methodology and relationship between these classes is also outlined in the section.

3.1. Experimental setup

In this experiment, three rhesus macaque monkeys (labelled Rs, Mk, Rj) participated in a **random target-pursuit task (RTP)**. Each monkey's arm rested on a two-joint exoskeletal robotic arm which corresponded their reaching movements to the movement of a visual cursor. Each monkey was trained to move the cursor to the visibly-presented target. Once the monkey passed the cursor over the target, the next target was presented at a random location on the surface. This was repeated without pause over the duration of an hour.

Each monkey had a 10×10 microelectrode array implanted in the arm area of its **primary motor cortex (MI)**. These were used to record electric field potentials over time, which can be filtered into two primary components:

- The **local field potentials (LFP)** are given by low-pass filtering and are thought to represent aggregate potentials from current entering the local area.

- The **action potentials** or **spike data** are given by high-pass filtering, corresponding to the output activity of individual neurons via **spiking** (rapid depolarization and repolarization of the cell's membrane potential).

For each monkey, we therefore received access to the following data (amplified, filtered, and pre-processed as outlined in [TKC+15]):

- Times when the target was reached (and the next target displayed)
- The LFPs at up to 96 of the 100 electrodes (**channels**) at a resolution of 1 ms
- The spike times for each of 59 to 115 neurons (**units**), with each neuron also corresponded to a channel
- The waveforms for every neuron spike (each recorded as 48 uniform time points over a duration of 1.6 ms).

3.2. Neuronal classifications

Unlike LFPs, which are inherently aggregate measures, spike data essentially encodes the behavior of individual neurons. Working across all the recorded neurons is what allows us to study **motor cortical population activity** as a whole. On the physiological side, there exist classes of neurons that are either functionally or physically distinct. Given the experimental apparatus, our hope is that we can distinguish classes based on both their individual spiking properties (firing rates over time, spike waveforms, etc.), as well as their behavior with respect to other neurons and even LFPs.

Here, we define and motivate some classifications we can take over our neuronal populations. Each class can then be analyzed via both conventional and topological data methods. If the classifications are chosen well, then the results can be given model-based neurological interpretations:

Narrow and wide neurons. The most apparent natural classification was made by examining the mean spike waveform widths (time difference between trough and peak of extracellular potential) of each neuron. In the case of our data, Takahashi et al. [TKC+15] showed that a bimodal Gaussian mixture model was the best fit for the distribution of waveforms for each monkey. Across monkeys, this leads us to define:

- **Narrow neurons** as those with spike widths of 0.2667 ms or less.
- **Wide neurons** as those with spike widths of 0.4000 ms or more.

Prior literature has performed similar classifications based on spike width and suggested that narrow neurons correspond to "local inhibitory neurons." However, [TKC+15] notes that the percentage of narrow neurons observed is much greater than the number of local inhibitory neurons in the neocortex (around 20%), and so there may be other types such as "excitatory pyramidal neurons" that may be included in this category.

We will primarily deal with subsets of narrow neurons, as almost all connected and oscillatory neurons are narrow as well, and because they have some agreed-upon interpretation. There are also other incidental differences between the two populations; for example, when the LFP beta power increased during a time window, narrow neurons also increased in power, while wide neurons decreased.

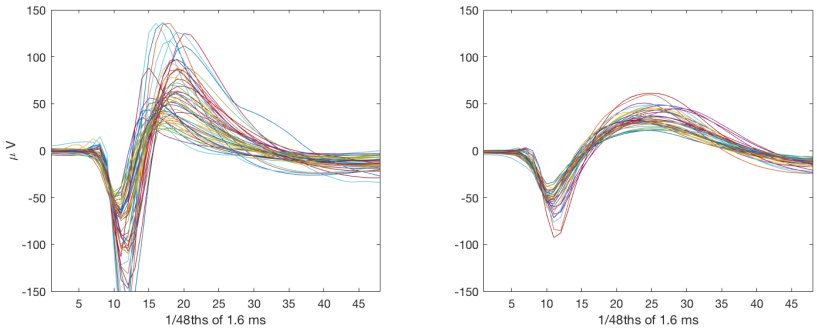


FIGURE 6. Mean spike waveforms for the narrow and wide neurons of Rs, respectively.

Connected neurons. A primary novel contribution of [TKC+15] was the use of a point-process generalization of Granger causality to give the spiking behavior of the motor cortical population a spatiotemporal interpretation. First, Granger causality allowed for the creation of a statistically-sound generative model where a neuron's spikes can be attributed to inhibitory or excitatory connections from other neurons and their spike behaviors. Then, via the unit-channel correspondence, we can infer the spatial orientation of each connection via the locations of the channel electrodes on the array (and thus along M1).

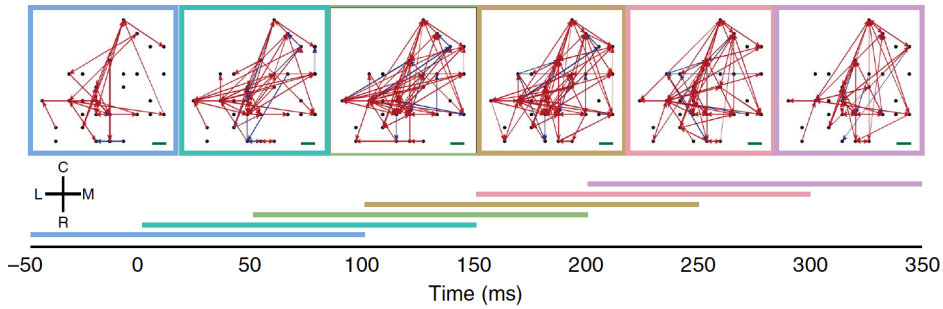


FIGURE 7. Inferred directed networks of connected neurons in Rs, corresponding to different (overlapping) time windows. (Image from [TKC+15].)

Furthermore, this analysis can be done over different relative time windows, which allows one to study how the functional connections between neurons evolve relative to the target hit time. For example, we informally see from Figure 7 that the number of effective connections peak around 100-250ms after the target becomes visible, where the connections are located as shown.

We say that the **connected neurons** for a relative time bin are those deemed connected by the Granger analysis. We consider this class since one might expect that neurons which are functionally connected have distinct properties from those which are not, or from the whole population itself, over that same partial window.

4. Topological analysis

In this section, we are motivated by the work of Singh et al. [SMI+08], who encoded the spiking rates of a cortical population into a point cloud data set, where the points corresponded to neuron spike counts per time window. We adapt their analysis in spirit and perform it on our populations of motor cortex neurons. In the same way, we derive point clouds corresponding to time windows, which we then filter for noise. Ultimately (up to a choice of persistent threshold), we get Betti signatures that we purport to describe the underlying topology of the neuron spikes.

4.1. Prior work

In the work of Singh et al., arrays of 100 electrodes were implanted in the V1 to directly detect time-dependent voltage signals, which were then processed using spike sorting to obtain firing times for N distinct neurons in the array's detection profile. They took data from two experimental situations: in the **spontaneous** case, the eyes of the animals were occluded, preventing direct visual stimulus and in the **evoked** case, monkeys were shown a sequence of movie clips. The spike-sorted data, which was a collection of all spike times for the N neurons, was first separated into 10 second segments, which were further sorted into 200 50ms bins. For each 50ms bin, they could then count the number of firing events of a given neuron within that bin. In each 10 second window, the five neurons with the highest firing rate were selected and for each bin, a vector in $\mathbb{R}^5$ was created whose components were the number of times each of the five neurons fired within the bin. The result is a point cloud of data consisting of 200 points in $\mathbb{R}^5$.

Beginning with these point clouds, a witness complex was constructed via the max-min procedure based on 35 landmark points. Looking at the distribution of the "topological signatures" across the 10-second data segments for both the spontaneous and evoked cases, the analysis suggests that the two main signatures dominate: a circle and a sphere. From this, it is inferred that the spontaneous and evoked cases exhibit similar topological distributions. Why would these topological structures exist within the visual cortex? The authors suggest the result is possibly a function of the following two facts, which are both experimentally supported: the primary visual cortex is tiled by maps of preferred orientation and spatial frequency and that extreme spatial frequencies tend to align with orientation pinwheels.

4.2. Activity and witness spaces

In general, given a distinguished class of K neurons and a distinguished time window relative to each hit time (e.g., [+150ms, +300ms)), one can construct a point cloud of data from their spikes in the following manner:

- Let N be the number of admissible time windows. That is, N corresponds to subset of target hits that we are considering. Relative to those hit times, we consider the relative time window.
- For the i-th time window, we take $\vec{x_i} \in \mathbb{R}^K$ such that its j-th coordinate is

$$\vec{x}_{i,j} = \ln((\#\text{ spikes of neuron } j) + 1)$$

This gives us N points, each with components in K dimensions. This follows the precedent set by [SMI+08], except they take $K = 5$ where the neurons

are chosen based on being the most active. Furthermore we take the logarithm of the spike count (plus 1 to maintain well-definedness) under the accepted model of neuron spikes within a window as Poisson-like, specifically where the spike counts of a neuron are logarithmic in a linear combination of its input neurons [TKC+15]. Intuitively, this is the idea that the existence of a 30th spike is not as indicative as that of a 3rd spike for a neuron over a time window. This also minimizes the impact of outliers.

Our operating assumption is that these points are sampled from some generative topological manifold embedded in $\mathbb{R}^K$, plus error. We shall call this implicit manifold the **activity space**. For example, in the case of Figure 1, the activity space would be the circle (or something homologous to it). It is this manifold whose topological properties we want to quantify up to homology. We do this by computing the persistent homology of witness complexes on the point cloud, with the belief that sufficiently stable properties (namely, Betti signatures that persist longer than a given **persistence threshold**) reflect properties of the underlying activity space.

For the analyses in the section, we used the following intermediate space to compute persistent homology:

- We take a witness complex of the point cloud, taking the first point to be our seed and then using max-min selection to produce a landmark set of 100 points.
- We directed the software to only keep track of the first three homology groups (H_0, H_1, H_2), for ease of computation.
- We compute homology over $\mathbb{Z}/2\mathbb{Z}$. While this cannot distinguish, e.g., between the Betti signatures of a Klein bottle and a torus, each Betti number will still encode the intuitive notion of number of independent cycle classes.

This project utilized JavaPlex, an open-source computational tool for computing persistent homology on point cloud data sets. For software download, see http://appliedtopology.github.io/javaplex/. For accompanying documentation, see https://github.com/appliedtopology/javaplex/wiki/Tutorial.

4.3. Persistence barcodes

Now that we have the notion of an activity space, we can vary the following variables and consider how the topology features change by observing the barcode of the derived witness complexes:

- The distinguished class of neurons
- The distinguished time window (relative to the hit time)

We parallel the analysis of Figure 7 by taking the relative time windows to be [-50ms, +100ms), [0ms, +150ms), etc. as well. Like in the second half of [TKC+15], or in the spontaneous versus evoked comparison of [SMI+08], our goal is to see *if and how the topology of the motor cortex evolves relative from when the next target becomes visible to the monkey.*

- The distinguished class of neurons below are the narrow neurons for monkey Rs.
- The admissible windows are subintervals from periods when the time between hits is not too short (e.g., < 300 ms) but also not too long (e.g., > 1 s).

Using these and the aforementioned parameters for constructing the witness complex gives Figure 8. It depicts the barcodes of the witness complex over the narrow neurons of Rs, as computed on shifting time windows relative to the hit time.

5. Conclusions

In this expository and preview paper, we motivated and established the mathematical theory behind persistent homology as a method of analysis that can capture topological properties of data. We then saw how problems in biology, and specifically in neuroscience, can be made amenable to topological analysis by converting the data into a discrete point cloud which encodes the relevant properties.

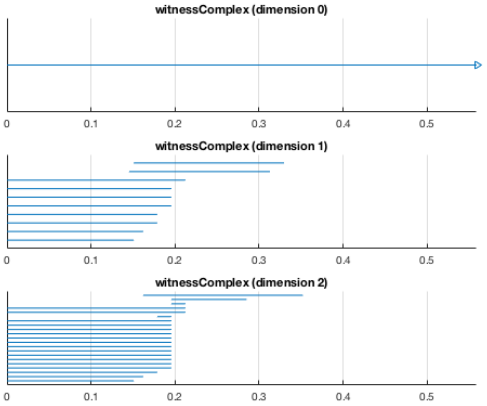
We introduced our experimental setup and data set for neuronal populations in the motor cortex, and converted activities into classes and point clouds of data, disambiguated by relative time windows, each with potentially different topological structure. In fact, in Figure 8 we see visually that there is a suggestive evolution of topological features that occur over the time windows, as the motor cortex switches into the active mode around 150ms to 300ms (the presence of longer/later 1 and 2-dimensional holes). This is all very informal; a significant amount of work remains to show that these differences are stable with respect to certain choices made (e.g., the choice of first point in the max-min procedure).

This paper shows that this line of inquiry can give results that conventional techniques might not, due to the distinctly deformation-independent nature of TDA. Our first future goal is make our analysis broader and more rigorous:

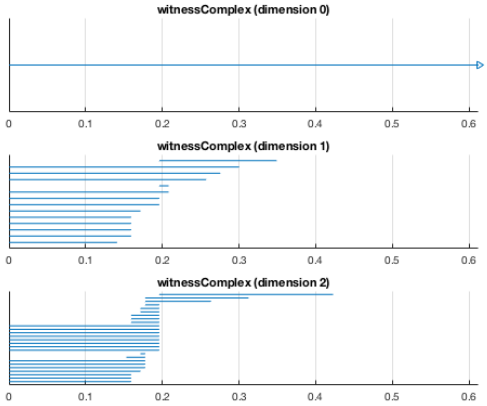
- Models must be validated; e.g., we want a mathematical/physiological justification versus an intuitive one of taking logarithms, max-min instead of weighted random, and other choices.
- Persistence should be quantified; e.g., how long does (1, 2, 1) persist as opposed to (1, 2, 2), etc.? If one takes different persistence thresholds, where only homology classes that last longer than 0.3, 0.4, etc., are taken to be meaningful, this may lead to different results.
- Statistical significance has not been tested; e.g., one should shuffle spike counts and seeing if the prevalence of certain Betti signatures changes. This tests the null hypothesis.
- Noise is an issue; perhaps some observations are outliers (imagine having noisy points in the middle of the circle in Figure 1; this would affect the inferred Betti signature by introducing an extra component). How do we detect and correct for this (e.g., only including points from regions of sufficient density)?
- We observed persistent homology in narrow neurons of Rs, but what about wide neurons? Connected neurons? Neurons across the three monkeys? Different choices of time windows?

Finally, the second future goal is to deduce a neurological basis for what our topological results actually mean in terms of brain functionality, as done by [SMI+08].

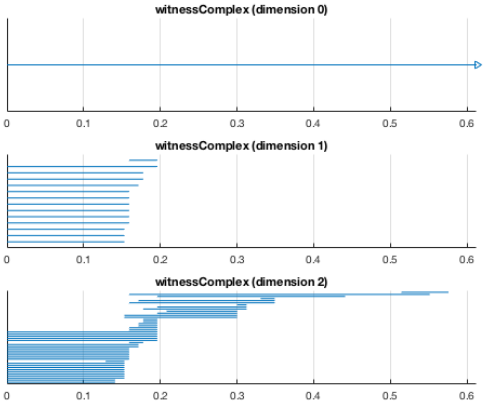
In sum, one should view persistent homology as yet another weapon in their data analysis arsenal, with all the possible choices and tradeoffs associated with it. We hope the reader is convinced of the viability of persistent homology as both a general analytical tool, and as a tool that can be novelly applied in this specific domain of studying neurons in the primary motor cortex.



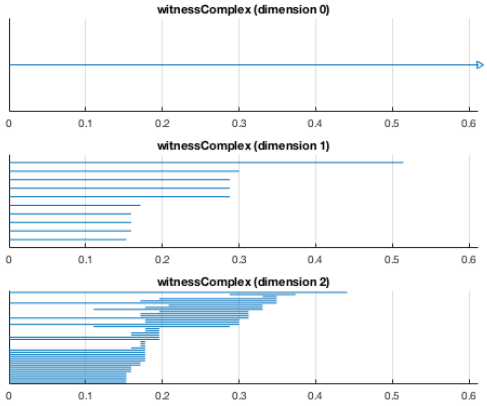
$[-50ms, 100ms)$



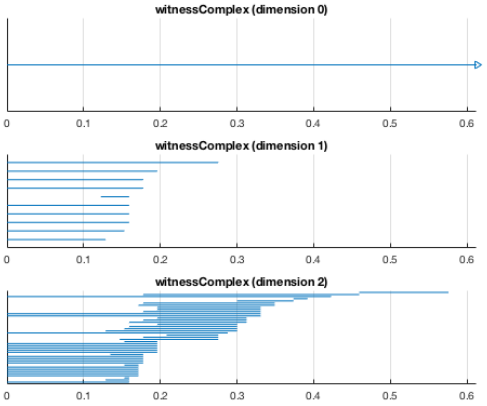
$[0ms, 150ms)$



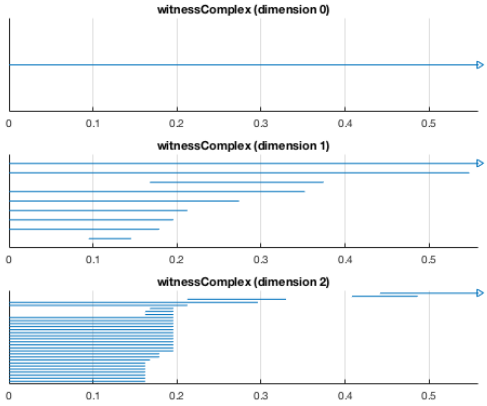
$[50ms, 200ms)$



$[100ms, 250ms)$



$[150ms, 300ms)$



$[200ms, 350ms)$

FIGURE 8. Persistence barcodes for the various time windows of narrow neuronal activity in Rs. Each line corresponds to a persisting element in the respective H_i. The x-axis is proportional to the filtration parameter ϵ.

Acknowledgements

The authors would like to thank Kazutaka Takahashi and the Hatsopolous Lab at the University of Chicago for sharing their expertise, advice, and data with us. We thank Jesse Wolfson for introducing us to persistent homology, and him and Professor Shmuel Weinberger for enlightening conversations regarding the relevant mathematical techniques and methods. Julian Salazar also thanks Professor Benson Farb for supporting his participation in the 2016 Mathematics REU at the University of Chicago. Emma West also thanks her REU mentor, Holly Mandel, for her support and thoughtful revisions.

References

[Car09] Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308, 2009.

[Dey13] Tamal K. Dey. Persistent homology. Computational Topology and Data Analysis Course Notes, Ohio State University (http://web.cse.ohio-state.edu/~tamaldey/course/CTDA/pers-homology.pdf), 2013.

[DSC04] Vin De Silva and Gunnar Carlsson. Topological Estimation Using Witness Complexes. In *Proceedings of the First Eurographics Conference on Point-Based Graphics*, SPBG'04, pages 157–166, Aire-la-Ville, Switzerland, Switzerland, 2004. Eurographics Association.

[Ede06] Herbert Edelsbrunner. Simplicial complexes & persistent homology. Computational Topology, Lecture Notes in Computer Sci- ence, pages 46-52, 128-134 (https://www.cs.duke.edu/courses/fall06/cps296.1/Lectures/sec-III-1.pdf), 2006.

[Hat02] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[HE00] A. Zomorodian H. Edelsbrunner, D. Letscher. *Topological persistence and simplification*. In IEEE Symposium on Foundations of Computer Science, 2000.

[Hen15] Richard Hennigan. A fast simplicial complex construction for computing the persistent homology of very large and high dimensional data sets. University of Massachusetts Lowell (http://www.cs.uml.edu/~rhenniga/main2.pdf), 2015.

[NLC11] M. Nicolau, A. J. Levine, and G. Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, Nov 2011.

[RM79] G.M. Reaven and G. Miller. An attempt to define the nature of chemical diabetes using a multidimensional analysis. *Diabetologica*, 16:17–24, 1979.

[Sal14] James Salvatore. Applying topology to data, part 2: More cech complexes, the vietoris-rips complexes, and clustering. Personal website (http://www.dyinglovegrape.com/math/topology_data_2.php), 2014.

[SMI+08] G. Singh, F. Memoli, T. Ishkhanov, G. Sapiro, G. Carlsson, and D. L. Ringach. Topological analysis of population activity in visual cortex. *Journal of Vision*, 8(8):11.1–1118, Jan 2008.

[TKC+15] Kazutaka Takahashi, Sanggyun Kim, Todd P. Coleman, Kevin A. Brown, Aaron J. Suminski, Matthew D. Best, and Nicholas G. Hatsopoulos. Large-scale spatiotemporal spike patterning consistent with wave propagation in motor cortex. *Nature Communications*, 6:7169, 2015.

[TKC+16] Kazutaka Takahashi, Sanggyun Kim, Todd P. Coleman, Kevin A. Brown, Aaron J. Suminski, Matthew D. Best, and Nicholas G. Hatsopoulos. Spatio-temporal patterning in primary motor cortex at movement onset. *Cerebral Cortex*, 2016.

[Wol16] Jesse Wolfson. Persistent homology. University of Chicago REU Lecture Series (https://jpwolfson.wordpress.com/2016-reu/), 2016.

[ZC05] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, Feb 2005.