Barcodes distinguishing morphology of neuronal tauopathy

David Beers,[1,*] Despoina Goniotaki,[2] Diane P. Hanger,[2] Alain Goriely,[1] and Heather A. Harrington[1]
[1]*Mathematical Institute, University of Oxford, Oxford, United Kingdom*
[2]*Institute of Psychiatry, Psychology, & Neuroscience, King's College, London, United Kingdom*

(Received 6 April 2023; accepted 17 August 2023; published 4 October 2023)

This paper is part of the Physical Review Research collection titled Physics of Neuroscience.

The geometry of neurons is known to be important for their functions. Hence, neurons are often classified by their morphology. Two recent methods, persistent homology and the topological morphology descriptor, assign a morphology descriptor called a barcode to a neuron equipped with a given function, such as the Euclidean distance from the root of the neuron. These barcodes can be converted into matrices called persistence images, which can then be averaged across groups. We show that when the defining function is the path length from the root, both the topological morphology descriptor and persistent homology are equivalent. We further show that persistence images arising from the path length procedure provide an interpretable summary of neuronal morphology. We introduce topological morphology functions, a class of functions similar to the Sholl functions, that can be recovered from the associated topological morphology descriptor. To demonstrate this topological approach, we compare healthy cortical and hippocampal mouse neurons to those affected by progressive tauopathy. We find a significant difference in the morphology of healthy neurons and those with a tauopathy at a postsymptomatic age. We use persistence images to conclude that the diseased group tends to have neurons with shorter branches as well as fewer branches far from the soma.

DOI: 10.1103/PhysRevResearch.5.043006

I. INTRODUCTION

Neurons are essential for cognitive function, and their activity is dependent on their morphology [1]. For example, when two simultaneous signals reach a neuron's dendrites, their mutual distance affects the combined signal that reaches the soma. In particular, nearby signals have a weaker total effect than those that are far apart [2]. Similarly, the length of the dendrites is known to affect the signal received by the soma, with signals that have traveled a long distance along the dendrites being weaker and more spread out than those that travel a short distance to the soma [3]. To further confound the study of neurons, it is known that neuronal morphology—even within the same animal [4]—can be highly heterogeneous, suggesting that different neurons have morphologies suited for different functions.

These observations have naturally led to attempts to classify large sets of neurons via their geometry. Computationally, neurons are typically represented by trees, collections of vertices in the two or three-dimensional Euclidean space connected by linear edges, with one such point denoting the location of the soma. Comparison of large classes of neurons can then be achieved by extracting numerical morphological descriptors, such as the number of bifurcations and branching angles, from these reconstructions [5–8]. However, these individual numerical descriptors cannot describe neuronal morphology in detail (e.g., the number of bifurcations cannot recover a neuron's tortuosity). Another approach consists in computing density maps that capture the distribution of neurites in a neuron [9–11]. One drawback of this approach is that if we have two classes of neurons with different density maps, it is difficult to accurately measure the difference between two density maps (i.e., minimize the integral of their difference over all possible rotations and translations).

A classical morphological descriptor is the *Sholl function* of a neuron [12], which assigns for every positive number r the number $s(r)$ of times a spherical shell of radius r intersects a given neuron. The Sholl function can be interpreted visually when plotted as a function of r. By taking averages of Sholl functions, researchers are able to visualize the average structure of large classes of neurons.

Given a neuron, represented mathematically as a rooted tree, and a function on its nodes, taken to be the Euclidean distance to the soma, two groups proposed independently the use of barcodes to compare morphologies [13,14]. Barcodes are mathematical objects used in topological data analysis as multiscale morphology descriptors. For instance, Li *et al.* [14] computed a barcode of neuronal data by applying *persistent homology* (PH) [15–17], a technique in computational mathematics for extracting multi-scale topological features from data. Li et al also show that the Sholl function of a neuron can be computed from the union of two particular barcodes arising from PH. In a separate work, Kanari *et al.* [13] developed an algorithm that takes a tree and function as

*david.beers@maths.ox.ac.uk

input, and calculates a PH inspired barcode, called a *topological morphology descriptor* (TMD). The TMDs of neurons are then converted to matrices called *unweighted persistence images* [18], to compare different classes of neurons [13,19]. Unweighted persistence images are representations of smooth two-dimensional functions called *unweighted persistence surfaces*. These unweighted persistence images are visualized as two-dimensional images with the intensity of the (i, j)th pixel representing the magnitude of the corresponding (i, j)th matrix entry.

Here, we build on the work of both groups. We compute TMDs, but we use the path distance, or *intrinsic distance* to the soma rather than the radial (Euclidean) distance from the soma as done in Ref. [13]. The intrinsic distance to the soma has already been suggested as a relevant choice of function [13,14] and briefly demonstrated on an example in Ref. [13, SI]; however, to the authors' knowledge it has not been thoroughly investigated for neuronal morphology.

A. Contributions

We show that different regions in the persistence images of TMDs generated with the path distance to the soma function can be interpreted to understand the morphology of large groups of neurons. Specifically, we discuss how *weighted* persistence images, derived from functions called *persistence surfaces*, record information about neuronal branches, including their length within a population of neurons as well as their distance from the soma. Based on this framework we have the following contributions.

(1) We prove that for a class of functions containing the path distance to the soma the barcodes of Refs. [13,14] are equivalent (Theorem 1).

(2) We define a new class of morphology descriptors, the *topological morphology functions*, and show that they can be recovered from barcodes derived using path length. Moreover, we show that topological morphology functions can be approximated from persistence surfaces via integration (Theorem 2).

(3) We apply these techniques to neurons from control mice and mice that model tauopathies, a form of neurodegenerative disorders in which tau protein forms deposits in the brain. This topological framework finds that neurons in mice with a postsymptomatic tauopathy are on average shorter than controls; furthermore, these neurons exhibit less branching far from the soma than controls.

The remainder of the paper is organized as follows. In Sec. II, we recall the how one constructs the TMD from a neuron equipped with a function, and how to represent a TMD as a persistence image. Section III describes the morphological information recorded by the TMD and its associated persistence images when a neuron is equipped with the path distance from the soma function. Further, this section exposes how the definition of topological morphology functions arises naturally from these TMDs, and shows how topological morphology functions may be approximately recovered from persistence images. Sections IV and V consist of an application of TMDs to study diseased mice. In the Appendix, we define the barcodes of [14], which are given by persistent homology. Here we also define some technical metrics on the space of persistence diagrams and prove theoretical statements made in Sec. III.

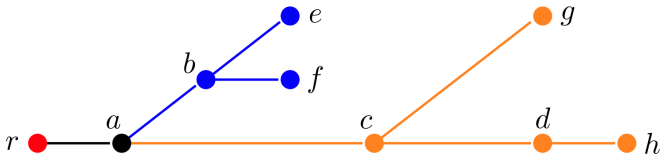

FIG. 1. A tree with root r. Nodes a, b and c are branch points, while nodes e, f, g, and h are leaves. Nodes r and d are neither. The two child branches of a are colored in blue and orange.

II. THEORY

A. The topological morphology descriptor (TMD)

Consider a neuron be represented by a tree T with nodes $N(T)$, and any function $f : N(T) \to [0, \infty)$ which returns a measure of distance from the soma. The *topological morphology descriptor* (TMD) of this neuron equipped with the function f is a collection of intervals in the real line that describes the branching pattern of the dendrites with respect to f [13]. To understand how the TMD is computed, we first introduce basic definitions regarding the tree T. In this paper, we consider only *finite trees*, i.e., trees with finitely many nodes. We also assume that all trees are equipped with embeddings in $\mathbb{R}^2$ or $\mathbb{R}^3$, sending edges to line segments, unless stated otherwise. Computationally, a neuron is represented by a tree T, with nodes associated to locations in $\mathbb{R}^3$, and with a distinguished node r that represents the soma, as shown in Fig. 2(a). We refer to the distinguished node as the *root* of T. From the root we induce an orientation of the edges of T. We orient any edge e incident to the r away from r. Inductively, if v is incident to an edge we have already oriented, we orient any edges incident to v and v', but not yet oriented, away from v towards v'. If, for a given v and v', in the resulting directed graph there is a directed edge e from v to v', denoted (v, v'), we say that v is the *parent* of v' and v' is a *child* of v. If a vertex v has three or more incident edges, we say that v is a *branch point*. If the root r has two or more incident edges then we say that r is a branch point. Similarly, if a vertex $v \neq r$ has exactly one incident edge, we say that v is a *leaf* of T. If r in a rooted tree has no incident edges, then r is a leaf. Suppose that v' is a child of v, i.e., there is a directed edge $e = (v, v')$. We can associate to v' the induced subtree T' generated by v', its children, its children's children, and so on. We say that the union $T' \cup \{e\}$ is a *child branch* of v. Examples of these definitions are shown in Fig. 1. In Ref. [13], the function f on T is chosen to be the Euclidean distance radial distance from the root r.

The methodology of Ref. [13] yields a local-to-global topological summary of the morphology of T using the function f as follows. At each branch point b in a tree T we inspect each child branch of b. We leave untouched the child branch that attains the greatest value of the function f on its leaves, and detach every other child branch of the branch point. It may be the case that at a given branch point, two or more different child branches attain the greatest value of the function f on their leaves. In such a case, we select, arbitrarily, one of these

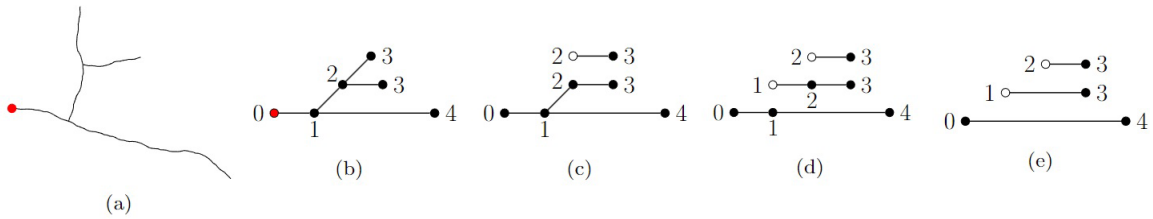

FIG. 2. Retrieving the barcode from a neuron. (a) A simple computer generated neuron, with its soma identified by a red node. (b) depicts the neuron represented as a tree with its root colored red. In (b), we also label each node by the values of a function f. (b) through (e) show the process of extracting a barcode from a rooted tree with labeled edges. In the transition from (b) to (c), we remove a branch at the branch point indicated by the number 2. Here, it does not matter which branch we remove since both branches attain the same maximal value of 3 on their leaves. In the transition between (c) and (d), we remove a branch at the branch point labeled 1. This time we do not have any choice of which branch to remove: one branch has a leaf with a value of 4, so we must remove the branch that only has a leaf with a value of 3. (e) shows the barcode we acquire after this process. Each interval corresponds to a branch in the original neuron, with its left endpoint given the f value at the branch point where the branch initiates, and the right endpoint given the f value at the leaf where the same branch terminates.

child branches to keep connected to the tree and we detach the other branches. This choice has no effect on the output of the algorithm. We then iterate over all branch points. Once this process has been completed at each branch point, what will remain is a collection of intervals, with endpoints given by their associated f values, which we denote by $\mathrm{TMD}(T, f)$. It may be the case that some interval appears more than one time in $\mathrm{TMD}(T, f)$. When this happens we record the number of times that the interval appears. We show this process applied to an example neuron in Fig. 2. Kanari *et al.* [13] choose a specific order in which their algorithm inspects the branch points; however, the chosen order does not change the resulting TMD, since applying the detaching procedure at one branch point has no effect on either the number of child branches any other branch point has, or the maximal leaf values of f on these child branches. In general, a multiset of intervals with endpoints indexed by real numbers, such as $\mathrm{TMD}(T, f)$, is referred to as a *barcode*. We summarize the procedure of generating $\mathrm{TMD}(T, f)$ in the following steps.

(1) Choose a branch point, and identify a child branch of that branch point that attains the greatest value of f on its leaves.

(2) Detach every child branch from this branch point except for the identified branch.

(3) If there is branch point in the resulting collection of trees, return to step 1.

(4) Label the endpoints of the resulting collection intervals with their associated f values.

Notably, there is a bijective correspondence between right endpoints of intervals and leaves of T. Hence the right endpoints must be closed. All but one of the left endpoints of $\mathrm{TMD}(T, f)$ are constructed by detaching a child branch from a branch point, and so must be open. The only interval with a closed left endpoint must be $[f(r), L]$, with L the maximum value of f on the leaves. To see this, consider the unique path along directed edges from r to a leaf l with $f(l) = L$. At each branch point on this path we can choose not to detach the child branch containing l during the TMD algorithm. Therefore the interval $[f(r), L]$ will be one of the remaining intervals after the TMD algorithm terminates. For an arbitrary function f, the intervals $[x, y]$ or $(x, y]$ of the $\mathrm{TMD}(T, f)$ are not always technically intervals in the real line, since it can be the case that $x > y$. However, if f is strictly increasing on paths along directed edges away from the root, branch points have lesser values of f than their descendants, and so left endpoints of intervals will always be less than right endpoints of intervals, provided T has more than one node.

B. Methods for analyzing TMDs

A barcode can be visualized as multisets of intervals (see Fig. 2). Barcodes can also be visualized as multisets of points in the real plane by sending each interval $[x, y]$ or $(x, y]$ to the point (x, y). We remark that this multiset records the *multiplicities* of points in the plane (i.e., the number of intervals with the same endpoints), and these multiplicities can be visualized (see e.g., Ref. [20, Fig. 2]). We call the disjoint union of this multiset of points with a copy of each (x, x) on the diagonal with infinite multiplicity a *persistence diagram*. These diagonal points are included for the technical purpose of defining distances between persistence diagrams, and are typically either not depicted visually or represented by a diagonal line. The bottleneck distance and the q-Wasserstein distances (both defined in the Appendix) are two frequently used distance functions between persistence diagrams which make use of these diagonal points. However for our purposes the diagonal points are not morphologically relevant. In Fig. 3, we show the example neuron from Fig. 2, its barcode, and its persistence diagram in panels (a), (b), and (c), respectively. The height of each point (x, y) above the line $y = x$, dashed in Fig. 3(c), is $y - x$, which is also the difference in f values of the endpoints of the corresponding interval in the barcode. By convention, we refer to x as the *birth* of the interval, y as the *death* of the interval, and the length $y - x$ as its *persistence*. The persistence diagram of $\mathrm{TMD}(T, f)$ is known to be stable for the bottleneck distance against perturbations of f, addition of short branches to T, and removal of short branches from T [13, SI, Theorem 1]. The same persistence diagram is also known to be stable for the 1-Wasserstein distance against a similar, somewhat more restrictive class of perturbations [21, Theorem 2].

Since the persistence of an interval is an important quantity, an intuitive transformation to apply is

$$(x, y) \mapsto (x, y - x). \tag{1}$$

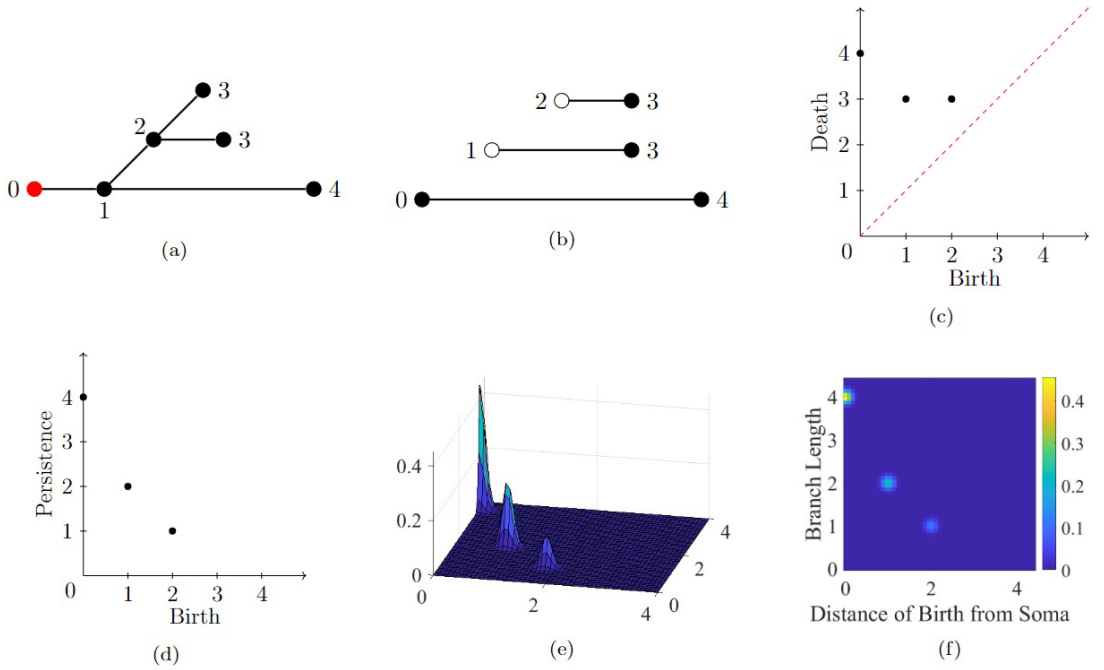

FIG. 3. Turning a tree into a persistence image. (a) shows an example tree with nodes labeled by their path length distance to the root, which is colored red. (b) and (c) show the barcode diagram and persistence diagram of this neuron, respectively. (d) shows the resulting transformed persistence diagram. Lastly, (d) and (e) show the resulting persistence surface and persistence image respectively. Note that the points with greater y values are weighted more heavily in the construction of (e) and (f).

In Fig. 3(d), we show the above transform applied to the persistence diagram in Fig. 3(c). This transformation leaves the birth value of each point unchanged, but replaces the death value of a point with its persistence. Note that this transformation is invertible with inverse $(x, y) \mapsto (x, x + y)$, and so does not lose any information from the persistence diagram.

Barcodes, diagrams, and their transformed counterparts can be convenient to visualize, but computing averages in the space of barcodes can be a rather complicated matter. Indeed, while averages (in particular Fréchet means) of persistence diagrams are known to exist [22, Theorem 12] and moreover can be computed [23], they are also known to be unstable and nonunique in general. These pathological traits of averages between barcodes are a result of the complex metric structure with which barcodes are equipped. For a simple example nonuniqueness and instability, we refer the reader to Fig. 4 of Ref. [24]. One solution proposed by Adams *et al.* [18] is to define a two-dimensional function that is the sum of Gaussian functions of a fixed standard deviation σ centered at each point in the transformed (birth, persistence) diagram. Each Gaussian function involved in this sum is given a weight equal to the height above the x axis of the corresponding point. Such a function associated to a barcode is called a *persistence surface* [18]. Figure 3(e) shows the persistence surface for the example neuron. As a result of the choice of weighting, the area under each such function is proportional to the sum of the lengths of intervals in TMD(T, f). In practice, persistence surfaces are often reduced to finite dimensional vectors called *persistence images*. Persistence images are matrices computed by fixing a grid of relevant values in the plane and defining the (i, j)th entry to be the integral of the persistence surface in question over the (i, j)th square in the grid[1] [18]. If the sizes of the squares in the grid are small relative to σ, then persistence images can be approximated by sampling points in the centers of each square. Most commonly, persistence images are displayed as heat maps, as in Fig. 3(f). An advantage of analyzing persistence images instead of barcodes or diagrams is the ability to efficiently compute differences and averages across of sets of neurons. The choice of standard deviation affects the resulting persistence image of a persistence diagram, and so should be regarded as an input parameter. While in Ref. [18, Sec. 6] the authors find experimentally that classification via persistence images is insensitive to the choice of σ, there are immediate consequences to choosing σ too high or too low. In the extreme limit $\sigma \to \infty$, the persistence surface approaches a constant function over the grid of interest. Meanwhile, if σ is chosen too low, it may be the case that resolution of the grid defining the persistence image is too coarse to approximate the persistence image by sampling a point in each grid square.

[1]Weighting the Gaussian functions is standard practice when generating a persistence image. As the authors of Ref. [13] skip the weighting step, they refer to persistence images of the TMDs they study as unweighted persistence images. The authors also do not transform their data prior to generating persistence images, since intervals can have negative persistence in an arbitrary TMD. Since we will restrict to barcodes with intervals of only positive persistence in the methods section, we transform the persistence diagram first, which is also standard.

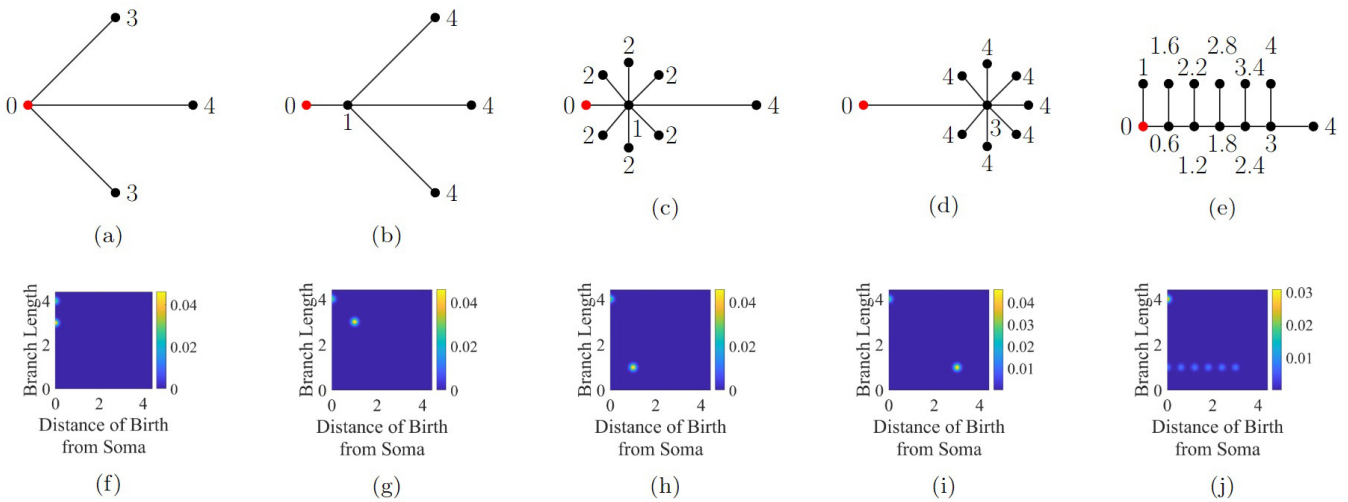

FIG. 4. Persistence images of toy neuronal trees. (Top row) Five toy neuronal trees with roots indicated by a red vertex. (Bottom row) Persistence images corresponding to the TMD(T, d) of the toy neuronal trees in the top row. All persistence images of these trees have a feature centered at (0,4), indicating that the furthest path length from the root (soma) of the toy neurons is 4 units. The remaining pixels with positive values correspond to the remaining branches of each toy neuron. (a) There are two additional branches of total length 6 that originate at the soma, which correspond to a feature centered at the coordinates (0,3) in (f). The relative pixel intensities reflect that these branches correspond to a total length of 6 while the longest branch only corresponds to a total length of 4. (b) Two shorter branches are a distance of 1 unit from the soma, which corresponds to a birth value of 1 and branch length 3 in (g). (c) Six branches of length 1 all initiate at a distance of 1 from the soma, which corresponds to birth and persistence value of 1. In (d) and (i), six branches of length 1 that initiate at distance 3 from the root causes an increase in pixel intensity near the coordinates (3,1). [(e) and (j)] Six branches of length 1 initiate at different distances from the root which correspond to positive pixels along the line with a y value of one, however, since they initiate at different distances from the root, the pixel intensity is dispersed across birth values.

While the distance of persistence images is not stable against perturbations to the bottleneck distance on the space of persistence diagrams, it is stable against perturbations to the 1-Wasserstein distance [18, Theorem 10] thanks to the weights applied to the Gaussian functions. A more subtle implication of changing σ than those detailed in the previous paragraph is that the stability guarantees of [18, Theorem 10] become weaker as σ approaches zero.

III. THEORETICAL RESULTS

A. Interpretation

We restrict our attention to barcodes given by TMD(T, d), where $d(v)$ is the intrinsic distance from r to v in T as a subset of $\mathbb{R}^3$. In other words, the function d returns the length of the path along the neuron from the point given by v to the soma. For example, d is used to obtain the TMD in Fig. 2. Since the TMD algorithm decomposes T into a collection of branches, the sum of the lengths of all the intervals gives the total branch length of the neuron. Further, the only closed interval in TMD(T, d) will be $[0, L]$, where L is the greatest value of d, since d attains all of its local maxima on leaves and $d(r) = 0$. All other intervals $(x, y]$ represent a branch in the barcode decomposition of T that initiates at an intrinsic distance x from the root and terminates at an intrinsic distance of y from the root. Therefore, the persistence of each interval in TMD(T, d) is exactly the length of the branch that the interval represents, which must be positive. Hence our choice of weights ensures the area under the persistence surface of TMD(T, d) is exactly the cumulative branch length of T.

A persistence surface of TMD(T, d) can be analyzed and interpreted back to its associated neuron in other ways as well. By identifying regions in the xy-plane where persistence surface $F(x, y)$ is large, we can read off the types of branches in TMD(T, f) that contribute to the total branch length. Indeed, regions where F is large with large y values correspond to longer branches, while regions where F is large with small y values correspond to the contribution of shorter branches. Similarly, regions where F is large with large x values correspond to branches that initiate close to the root. By contrast, regions where F is large with small x values correspond to branches that initiate far from the root, in the sense that the path from the root to the branch point at which they initiate is long. Note that the persistence image will always have a Gaussian contribution of weight L centered at $(0, L)$, since TMD(T, d) must contain the interval $[0, L]$. Since persistence images are discrete approximations of persistence surfaces, they inherit similar interpretations. These concepts are illustrated with artificial neurons, each with the same total branch length, in Fig. 4. In Fig. 5, we show how persistence images can summarize the traits of real morphologically distinct neurons. As mentioned, an advantage of persistence images over persistence diagrams is that it is straightforward to compute and visualize the average persistence image of a collection of neurons. From average persistence images, we can identify the types of branches that tend to appear in an entire class of neurons.

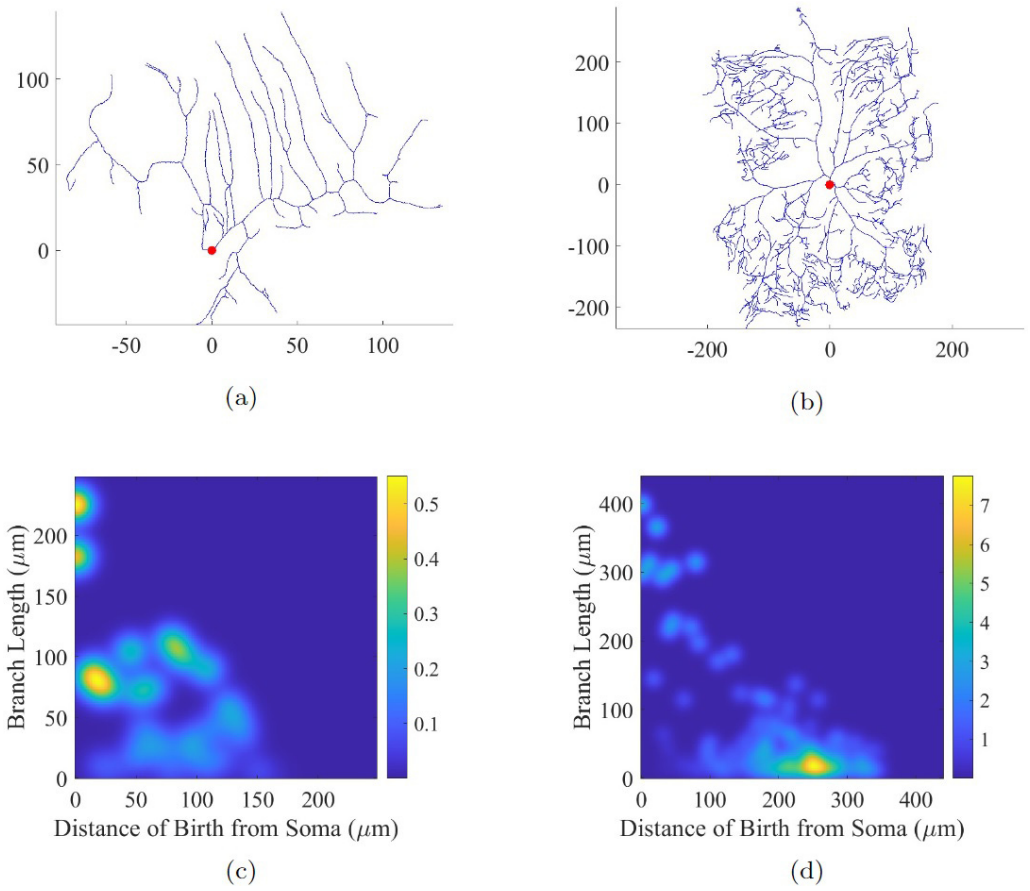

FIG. 5. Persistence images of *Drosophila* neurons. (a) A class I neuron characterized by few main branches and many long secondary branches growing perpendicularly to the main branch. (b) An area-covering class IV sensory neuron with many small branches far from the root. In both (a) and (b), the soma is indicated by a red node. Below each in (c) and (d) are the corresponding persistence images. There are several bright regions substantially above the x axis in (c) resulting from the many long secondary branches in (a). The brightest region of (d) is far to the right and close to the x axis since many of the branches of (b) are far from the root and short. The digital reconstructions of these neurons from Ref. [25] are freely available, see Ref. [26]. These neurons are named 02-16-09-Class1-B40X and 02-16-09-ClassIV on the site, respectively.

B. The TMD versus persistent homology

As previously mentioned, Li *et al.* [14] have developed a method distinct from the TMD which also takes a tree T equipped with a function f and returns a barcode, which we denote by $\mathrm{EPH}_0(T, f)$. We show that these methods are highly related.

Theorem 1. Let T be a finite rooted tree with root r, and f be a function which is strictly increasing along the directed edges induced by r. We have that

$$\mathrm{TMD}(T, f) = -\mathrm{EPH}_0(T, -f).$$

Here, the negative of a barcode denotes the operation which takes the negative of interval endpoints and reverses their order. In particular, d is a function satisfying the requirements of this theorem, and so calculation of $\mathrm{TMD}(T, d)$ can be recast as a calculation using the methods of Li *et al.* [14]. For the sake of succinctness, we defer the technical details of calculating $\mathrm{EPH}_0(T, f)$ and a proof of Theorem 1 to the Appendix. When f does not satisfy the hypothesis of the theorem, there is no guarantee that these two methods are related. For example, when f is radial distance, the existence of branches that grow towards the soma may cause the equality of Theorem 1 to not hold.

C. Proposing topological morphology functions

It was shown in Ref. [14] that a neuron's Sholl function can be reconstructed from a barcode obtained via an alternate methodology. We show that a similar function can be recovered from the barcode $\mathrm{TMD}(T, d)$. With the function d we lose information about the radial distance of branches from the soma necessary for the construction of Sholl functions. However, we can construct another family of functions which similarly describe the branching morphology of a given neuron. Explicitly, we define the *topological morphology function* p of a neuron to be the function which associates to each positive number t the number of points on the neuron which have an intrinsic distance of t to the soma. This is an example of a Sholl descriptor, a kind of morphological descriptor defined and studied in [27]. The value $p(t)$ is the number of intervals in $\mathrm{TMD}(T, d)$ containing t, since this is the number of branches in the TMD decomposition of T containing points

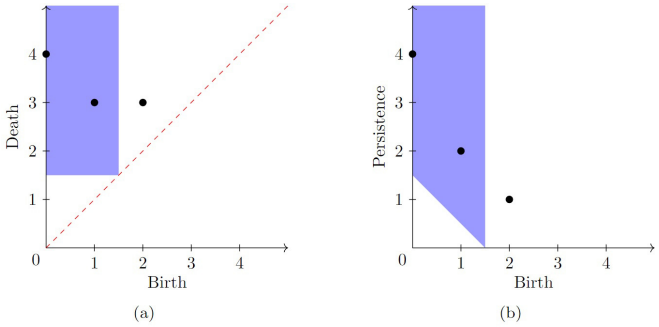

FIG. 6. Recovering $p(1.5)$ from a standard and transformed persistence diagram. On the left, $p(1.5)$ is recovered from the persistence diagram by counting the number of points above and to the left of the point (1.5, 1.5). This region is shaded blue. On the right, $p(1.5)$ is recovered from the corresponding transformed persistence diagram by counting the number of points in R_t, shaded blue. In both cases, the number of points in the blue region is 2, so $p(1.5) = 2$.

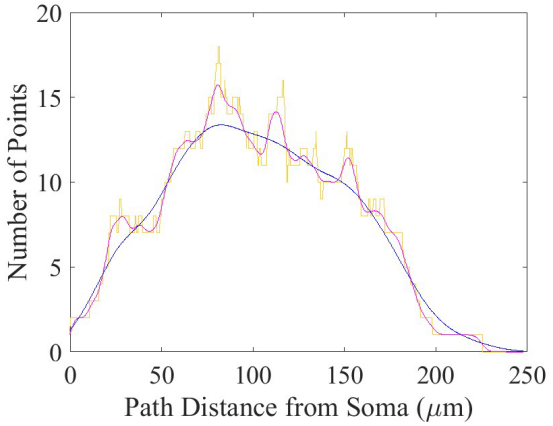

FIG. 7. The topological morphology function (orange) for the neuron in Fig. 5(a) along with the approximate topological morphology function obtained by numerical integration of the persistence image for $\sigma = 10$ (blue) and 2 μm (purple).

a distance of t from the soma. We call a number t *generic* if it is not an endpoint of an interval in $\mathrm{TMD}(T, d)$. For generic t, the number $p(t)$ is easily obtained from both persistence diagrams and transformed persistence diagrams, as shown in Fig. 6. In a persistence diagram, $p(t)$ for a generic t is the number of points above and to the left of (t, t). In a transformed persistence diagram, $p(t)$ is the number of points in the region R_t, given by the equations

$$x < t, \quad y > t - x, \tag{2}$$

for generic t. The reason why we can only deduce $p(t)$ from persistence diagrams for generic t is that persistence diagrams do not retain information about the openness or closedness of the intervals in their corresponding barcode.

The following theorem shows that the topological morphology function p of a neuron represented by T can be approximated from persistence surfaces of $\mathrm{TMD}(T, d)$ via integration.

Theorem 2. Let p be the topological morphology function associated to a neuron represented by a tree T. Let $d : N(T) \to [0, \infty)$ be the intrinsic distance to the soma. Let $F_\sigma(x, y)$ be the persistence surface corresponding to $\mathrm{TMD}(T, d)$ constructed with Gaussian functions of standard deviation σ. For any generic positive number t,

$$p(t) = \lim_{\sigma \to 0} \int_{R_t} \frac{F_\sigma(x, y)}{y} \, dx \, dy. \tag{3}$$

Further the integral in the above expression converges and is an infinitely differentiable function of t for all $\sigma > 0$.

This theorem shows that we may use the persistence surface of a barcode to retrieve a smooth approximation of the topological morphology function, which improves in accuracy as the standard deviation σ used to generate the persistence surface approaches zero. We show on an example in Fig. 7 that approximation of the integral in Theorem 2 by numerical integration over a persistence image can also recover an approximation of the function p. Note that if every branch of a neuron grows in a straight line exactly radially outward from the soma, then the topological morphology function of the neuron is equal to the neuron's Sholl function s. In situations where this is a reasonable approximation, we expect that p and s are similar. Under such an assumption regarding the direction of neurite outgrowths, this theorem shows that we can use a persistence image to approximate the Sholl function s smoothly.

We defer the proof of Theorem 2 to the Appendix, where we also prove an analogous result for the barcodes of Li *et al.* [14].

IV. METHODS

A. Data

We study Tau35 mice, a transgenic mouse line which has been used as a model to reproduce biological and cognitive features of human tauopathies [28]. In particular, we analyze topologically a set of 316 digitally reconstructed mouse neurons (as `.DAT` files), 170 wild type (WT) and 146 Tau35 mouse variant, to study the effect of tauopathy on morphology. These reconstructed neurons are further divided into groups by age (e.g., presymptomatic 4 month neurons or postsymptomatic 10–12 month neurons) and type (e.g., cortical or hippocampal neurons). All neurons in the dataset were grown in live mice and measured via a staining technique. We refer the reader to Sec. A of the Appendix for the details of the data acquisition. A portion of this data is recorded via its projection onto the xy plane, while remainder of the data is recorded as a shape in three-dimensional space.

B. Preprocessing

We convert the data from `.DAT` to `.swc` format using NLMorphologyConverter. In `.swc` format, neurons are represented by a point cloud of points on the cell, with adjacent points on neurites indicated. Each `.swc` representation also features a collection of points representing the boundary of the soma. Since our methods do not take into account the morphology of the soma, we further preprocess the data by contracting these boundary points to a single point at their center of mass. By visual inspection, we find that the converted `.swc` converted files often attach neurites to the soma

TABLE I. The counts of neurons of different types in the dataset. The rightmost four columns denote the percentage of random regroupings which had an L^1 distance of group averages less than the L^1 distances of the WT and Tau35 averages for different values of the parameter σ used to define persistence images.

Age(Months)	Region	WT Count	Tau35 Count	Percent $\sigma = 5\,\mu m$	Percent $\sigma = 10\,\mu m$	Percent $\sigma = 15\,\mu m$	Percent $\sigma = 20\,\mu m$
10-12	Cortex	64	43	99.57	99.06	98.65	97.94
10-12	Hippocampus	38	49	97.38	99.68	99.76	99.79
4	Cortex	38	33	99.08	98.54	96.98	93.55
4	Hippocampus	30	21	85.20	94.87	94.46	91.51

at inaccurate locations. To correct for this, we further preprocess the data by removing edges between each neurite and the soma, and adding a new connection between the point representing the soma and each neurite, at the point on the neurite that is closest to the contracted soma point.

C. Computations

Since a portion of the data is projected into two dimensions, we use projected path length to the soma as a proxy for path length to the soma. We define the projected distance between any two adjacent points in a `.swc` representation of a neuron to be their mutual distance once projected in the xy plane. The projected path length between a point z and the soma is then defined to be the sum of projected distances between pairs of adjacent points along the shortest path between z and the point representing the soma. Letting $\widetilde{d}$ denote projected path length to the soma, we compute persistence images of TMD$(T, \widetilde{d})$ associated to each neuron in the data set and take in-group averages. Each persistence image is generated from the persistence surfaces[2] of Gaussian functions with a standard deviation of 5 μm. We hypothesize that the difference between group averages of WT and Tau35 neurons of the same age and type is significant. Since we do not know the distribution underlying the persistence images of our data, we opt to perform a nonparametric test of our hypothesis. Therefore we compute the L^1 distance between average persistence images of the Tau35 and WT neurons in question. We then compare this L^1 distance to the L^1 distances for averages of 10000 randomized regroupings of the data. We then repeat the above computations with the value of the parameter σ changed to 10, 15, and 20 μm to study the sensitivity of our analysis to the parameter σ.

V. RESULTS ON TAUOPATHY VS WT NEURONS

To test the hypothesis that WT neurons and tauopathy neurons are on average morphologically distinct, we average the persistence images of each group of neurons and perform a randomization test as explained in the Methods. We show the results of our randomization tests in Table I, where the rightmost column is the percent of times the randomized L^1 distance was less than or equal to the same distance between the average Tau35 and WT neurons. For each of the four randomization tests we choose to accept the between-average difference as significant if less than 5/4 percent of the regroupings had a greater between-average distance. We divide by 4 to account for the multiple comparisons problem, since we perform four tests for each value of σ . Under this criterion, the between-average difference of the persistence images of Tau35 and WT neurons was significant for 3/4 tests on 10–12 month hippocampal neurons, 2/4 tests on 10–12 month cortical neurons, and 1/4 tests on 4 month cortical neurons. In Fig. 8, we show the average persistence images (for $\sigma = 10$ μm) of the Tau35 and WT neurons of each 10–12 month group with their in-group differences.

We analyzed in-group averages of persistence images to identify features which distinguish WT neurons from Tau35 neurons on average at 10–12 months. For both hippocampal and cortical neurons in this age range [Figs. 8(c) and 8(f), respectively], we observe that the average Tau35 persistence images have a greater intensity towards the origin, while WT persistence images have greater pixel intensity for larger x and y values. From these observations we deduce that branches tend to not only be longer in WT neurons, but also often initiate further along the neuron than in Tau35 neurons. These findings are in accord with previous work associating presence of toxic tau with the inhibited ability of neurons to maintain long range connections in tau35 neurons [29] and neurodegenerative disease in general [30–34].

In Fig. 9, we plot the associated Sholl and topological morphology functions of the Tau35 and WT 10–12 month cortical and hippocampal neurons. We observe from both the Sholl and persistence functions that there are less branches far from the root in Tau35 neurons than WT neurons. The observation that the topological morphology and Sholl functions are increasing at greater distances from the soma for WT neurons, suggests that branching tends to occur at greater distances from the root in WT neurons. The analysis of the persistence images, which capture the joint distribution between branch initiation and branch termination confirms this conclusion.

VI. CONCLUSION

Barcodes and persistence images are powerful tools for the analysis of tree-like geometries in biological systems, such

[2]For computations we generated persistence images with 231×231 entries. The entries of these persistence images evenly cover the region $[-0.15L_{\max}, L_{\max}]^2 \subseteq \mathbb{R}^2$, where $L_{\max}$ is 1.1 times the maxiumim in-group projected branch length, across Tau35 and WT neurons. Note that the persistence images we use for computation sample values outside the first quadrant to assure that relevant information of points on the associated persistence diagram near the x or y axis is captured.

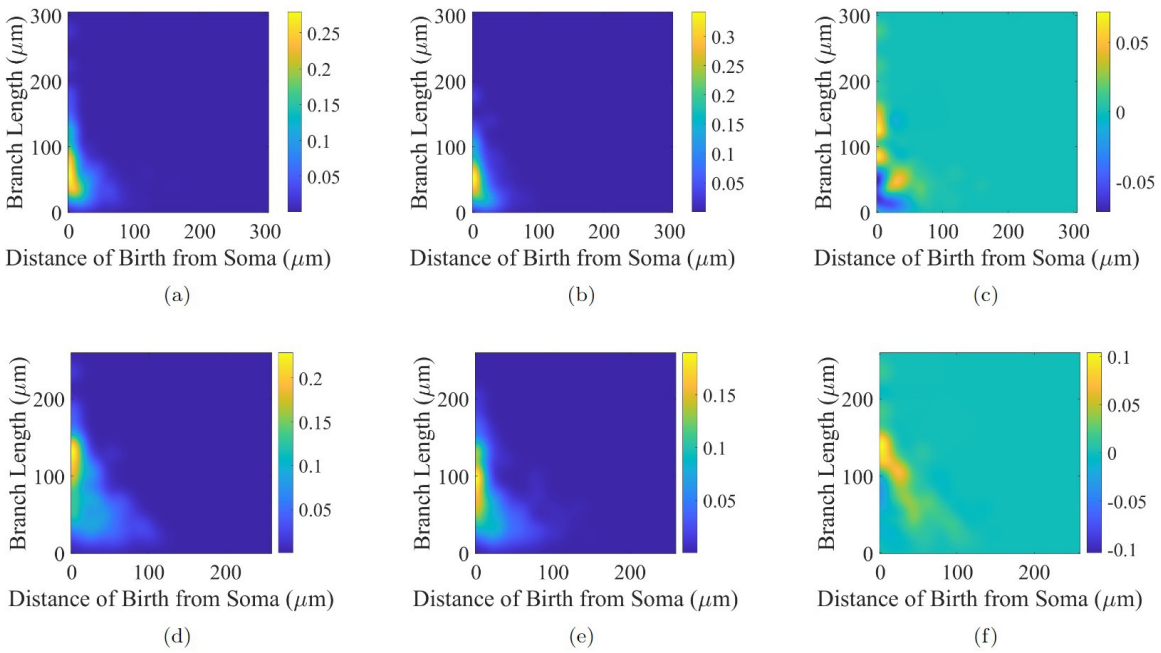

FIG. 8. In-group average persistence images and the difference between Tau and WT groups for 10-12 month cortical (first row) and hippocampal (second row) neurons. (First column) average WT persistence images, (second column) average Tau35 persistence images, (third column) difference between WT and Tau35 average persistence images.

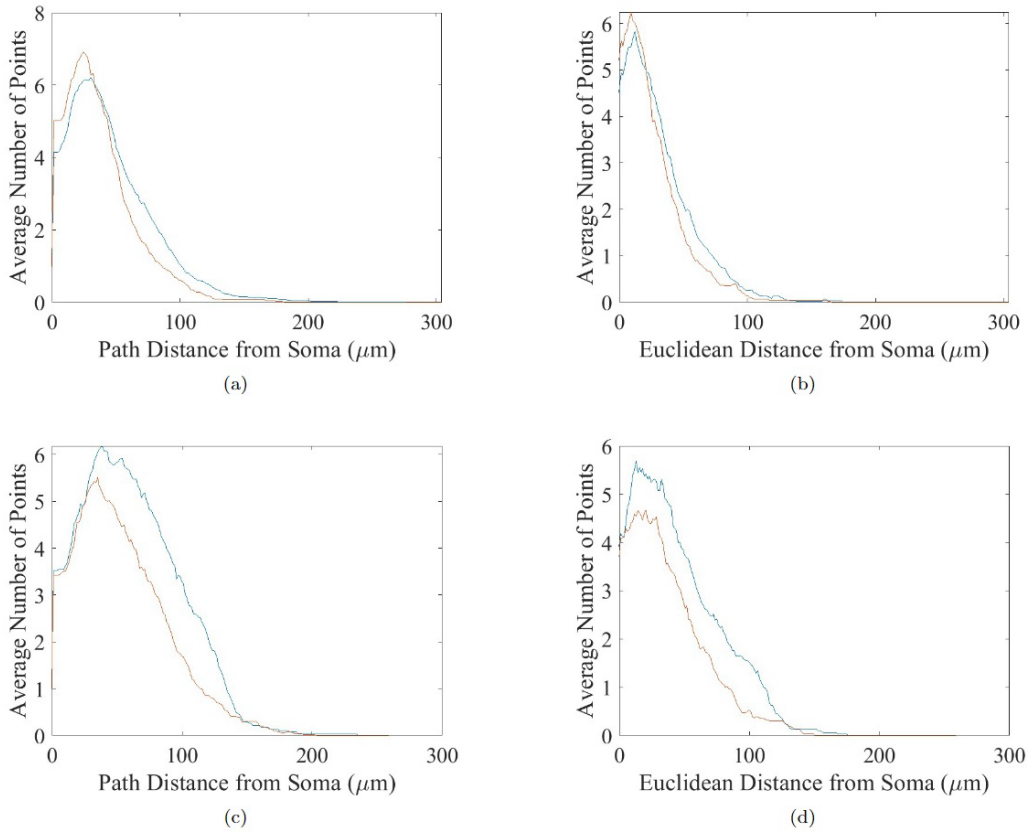

FIG. 9. Average (left column) topological morphology and (right column) Sholl functions for (top row) 10–12 month cortical and (bottom row) 10–12 month hippocampal neurons of both WT (blue) and Tau35 (orange) classes.

as neurons. An advantage of persistence images is that they can be used to study statistically groups of neurons. Here, we have further shown that persistence images based on a notion of path length capture key features of neuronal morphologies and can be used to discriminate statistically distinct classes of neurons, however in contrast with the work of Ref. [18], we find that the choice of the parameter σ plays a decisive role in our analysis. Unlike the Sholl functions, persistence images are able to describe the joint relationship between branch initiation and termination in neurons. Once two groups of neurons have been shown to be statistically distinct, the difference between the two groups can be determined through visual inspection of their respective averaged persistence images and their differences. When applied to neurons in healthy and diseased groups, we have shown that the quantitative features of the branches of a neuron or a group of neurons can be easily interpreted from our persistence images.

The natural choice of path length from the soma as an input to the topological morphological descriptor leads to the definition of a topological morphology function similar to the Sholl function. Smooth approximations of this function can be obtained directly from persistence images and in the particular case when branches mostly grow away from the soma, this topological morphology function produces an approximation of the Sholl function. By presenting an alternative function for the TMD, we are able to precisely connect the TMD algorithm to persistent homology and contribute an interpretable descriptor of intrinsic neuronal morphology that complements the toolkit of neuronal analysis.

ACKNOWLEDGMENTS

D.B. thanks Jacob Leygonie for helpful discussions. This research was funded in whole, or in part, by UKRI EP/R018472/1. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. D.B. is grateful for the support of the Mathematical Institute of Oxford. H.A.H. gratefully acknowledges funding from EPSRC EP/R005125/1 and EP/T001968/1, the Royal Society RGF\EA\201074 and UF150238. The support for A.G. by the *Engineering and Physical Sciences Research Council* of Great Britain under research grant EP/R020205/1 is gratefully acknowledged. This work was supported by grants from the Alzheimers Society and the Alzheimers Research UK Kings College London Network Centre to D.G. and D.P.H..

APPENDIX

1. Data Acquisition

a. Tissue harvest and processing

Tau35 mice were generated by targeted knock-in to the *Hprt* locus under the control of the human tau promoter as described previously [28]. Mice, Tau35 and wild-type (WT) controls, were sacrificed at two time-points, presymptomatic (four month old) and post-symptomatic (10–12 month old), using terminal anaesthesia, perfused with 1x PBS and post-fixed in 4% (w/v) PFA overnight at 4 °C. Brain sections (200 μm) were prepared using a VT1000 S Vibrating blade microtome (Leica Biosystems) and stored free-floating in cryoprotectant (30% (v/v) ethylene glycol, 15% (w/v) sucrose in PBS) at −20 °C. All procedures were carried out in accordance with the Animals (Scientific Procedures) Act, 1986, following approval by the local ethical review committee.

b. Tissue staining and image acquisition

To quantify dendritic morphology, Golgi-stained neurons [35] were imaged using a 60x objective (NA=1.4) on a Nikon microscope. Using the live acquisition feature images were collected at a depth starting at 20 μm below the surface of the specimen. Z-stack images (30–90 μm total on the Z axis; z-stack step size=0.3 μm; 90–270 images per stack) were acquired. Each image stack was extracted using NIS-Elements (Nikon) software and imported to Neurolucida (MBF Bioscience) software for analysis. Neurolucida was used to identify cell bodies and their outgrowth for each of the sections in the z-stack for stacks below 50 μm. For stacks above 50 μm, due to the image size, an average intensity projection image was generated by calculating the average intensity (AIP) values of each pixel along the z axis for all the layers and combining them into the 2D AIP image. The resulting AIP images were then used to identify cell bodies and outgrowth with the NEUROLUCIDA software.

2. Zero-dimensional persistent homology for graphs

This section of the Appendix constitutes a brief outline of the concepts from persistent homology that are necessary for other sections of the Appendix. We refer the interested reader to Refs. [36] and [16, Chap. VII] for a more complete overview.

Suppose we have a tree T and a map $f : N(T) \to \mathbb{R}$. We can define the *sublevel sets* T_t to be the subgraphs of T, generated by all of the nodes v satisfying $f(v) \leqslant t$. Any function on T has only finitely many distinct sublevel sets T_t, each of which has a lowest corresponding t value t_i. Hence we have a nested inclusion of graphs

$$\varnothing \subseteq T_{t_1} \subseteq T_{t_2} \subseteq \ldots \subseteq T_{t_n} = T,$$

which we can also write as

$$\varnothing \to T_{t_1} \to T_{t_2} \to \ldots \to T_{t_n} = T, \tag{A1}$$

with each arrow denoting the inclusion map. This sequence of graphs is called the *sublevel set filtration of* f. We depict an example sublevel set filtration in Fig. 10.

Each forest T_{t_i} has an associated vector space $C_0(T_{t_i})$ given by the $\mathbb{R}$-span of the nodes of T_{t_i}. The inclusion maps in the sequence of T_{t_i} then induce linear maps in the sequence

$$\{0\} \to C_0(T_{t_1}) \to C_0(T_{t_2}) \to \ldots \to C_0(T_{t_n}) = C_0(T).$$

We can quotient each vector space $C_0(T_{t_i})$ via the relation $v \sim v'$ whenever v and v' are in the same connected component in $C_0(T_{t_i})$. We call the resulting vector spaces $H_0(T_t)$, the *zero-dimensional homology* of T_t. Hence, $H_0(T_t)$ has dimension equal to the number of trees in the forest T_t. If v and v' are both in the same connected component of T_t, it follows that they are also both in the same connected component of T_a for

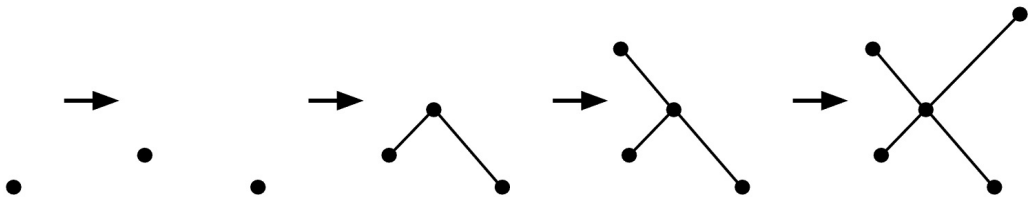

FIG. 10. The sublevel set filtration of the height function on the nodes of the rightmost tree.

any $a \geqslant t$. Therefore the inclusion maps above induce maps between homology groups, giving us the sequence

$$\{0\} \to H_0(T_{t_1}) \to H_0(T_{t_2}) \to \cdots \to H_0(T_{t_n}) = H_0(T).$$

This sequence connected by the linear maps induced by inclusion is called the *dimension zero persistence module* of T with respect to f. We would like to decompose this sequence of vector spaces into simpler sequences of form

$$\{0\} \to \cdots \to \{0\} \to \mathbb{R} \to \cdots \to \mathbb{R} \to \{0\} \to \cdots \to \{0\}$$

with identity maps between the spaces $\mathbb{R}$ and zero maps elsewhere. We refer to such a sequence of vector spaces with $\mathbb{R}$ first appearing at the location for t_i and last appearing at the location of t_j as $I(t_i, t_{j+1})$, and we call this type of sequence an *interval module*. For when $j = n$ in this definition, we define $t_{n+1} = \infty$. It was shown in [37] that if $\mathcal{M}$ is a zero-dimensional persistence module of T with respect to f, then $\mathcal{M}$ decomposes uniquely into a direct sum of interval modules

$$\bigoplus_{k=1}^{K} I(t_{i(k)}, t_{j(k)}).$$

From this decomposition, we associate a barcode, a collection of intervals in the real line, by the mapping

$$\bigoplus_{k=1}^{K} I(t_{i(k)}, t_{j(k)}) \mapsto \bigsqcup_{k=1}^{K} [t_{i(k)}, t_{j(k)+1}).$$

We refer to this barcode as the *persistent homology* of T and f, or $\mathrm{PH}(T, f)$. It follows from this decomposition that there are exactly as many infinite intervals, intervals with right endpoint $t_{n+1} = \infty$, in $\mathrm{PH}(T, f)$ as there are connected components in T. Since we assume T to be connected, exactly one of these intervals is infinite. Sometimes it is useful to make the resulting barcode only consist of bounded intervals, and we can achieve this, for example, by sending the one instance of ∞ to $\max(f)$ and replacing the open endpoint with a closed endpoint. We call the resulting barcode $\mathrm{EPH}_0(T, f)$. Readers familiar with extended persistence [38] will note that this is the zero-dimensional extended persistence barcode of T, justifying the notation. We show the barcodes $\mathrm{PH}(f)$ and $\mathrm{EPH}_0(f)$ of the sublevel set filtration from Fig. 10 in Fig. 11.

In Ref. [14], barcodes of bounded intervals for a tree T and a function f are obtained by first computing $\mathrm{PH}(T, f)$ and then transforming the result to the barcode of bounded intervals $\mathrm{EPH}_0(T, f)$ as described above. In this setting, $\mathrm{PH}(T, f)$ is computed by the following procedure under the assumption of certain genericity conditions. It is well known that there is a bijection between intervals in $\mathrm{PH}(T, f)$ and the local minima of f. The local minima are the subtrees of T on which f takes exactly one value which is less than its value on all neighboring nodes. If f is generic then all local minima of f are nodes. The values $f(v)$ for minima are the values of the left endpoints of their associated intervals, since they correspond to t values where their associated connected components first appear. Suppose that at time t_i in the sequence given by Eq. (A1), two or more connected components $\{C_l\}_{l=1}^{N}$ of $T_{t_{i-1}}$ merge. Additionally, suppose there exists a unique global minimum of f restricted to each C_l, denoted v_l, and one of the v_l has the lowest f value, without loss of generality $f(v_l) < f(v_1)$ for $l \geqslant 2$. In this generic case, then the Elder Rule [39, Theorem 4.4] determines that the right endpoint of the interval corresponding to v_l to be $f(t_l)$ for $l \geqslant 2$, giving rise to the interval $[f(v_l), f(t_i))$. If our assumptions hold whenever connected components merge, we can compute the persistent homology by this procedure. Indeed, after applying this process to every merging of connected components in Eq. (A1), every local minimum whose corresponding right endpoint has yet to be assigned must be the global minimum of its connected component in T. Since T is connected, there must only be one such local minimum. Hence, the value of this local minimum must be paired with an infinite right endpoint, as $H_0(T)$ is a one-dimensional vector space.

We can still compute $\mathrm{PH}(T, f)$ in a similar fashion when f is not generic. If $M_1, \cdots, M_k$ are the local minima of f, then we have k intervals in $\mathrm{PH}(T, f)$, each with left endpoint

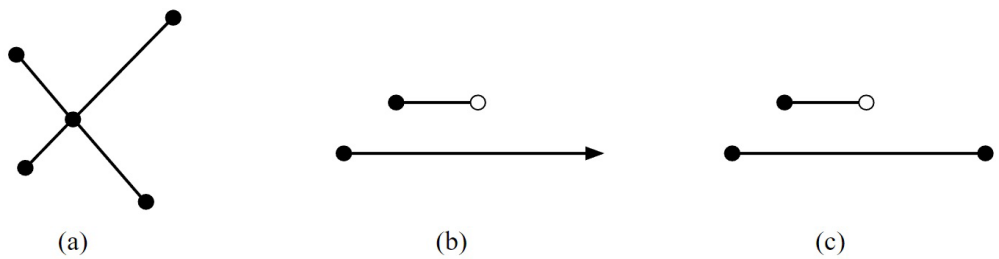

FIG. 11. (a) The example tree T from Fig. 10. (b) The barcode $\mathrm{PH}(T, f)$, where f is the height function on the nodes. (c) The barcode $\mathrm{EPH}_0(T, f)$.

$f(M_k)$. This time inducting on i, suppose again that at time t_i, connected components $\{C_l\}_{l=1}^N$ of $T_{t_{i-1}}$ merge, but by induction suppose exactly one of the local minima in each C_l has not yet been assigned a right endpoint. Indeed, there must be exactly one such local minimum for the base case $i = 1$ where every C_i must be a local minimum. For general i, refer to this distinguished local minimum in C_l by M_{n_l}. By reordering indices, we can have that $f(M_{n_1}), \cdots, f(M_{n_N})$ is weakly increasing. The Elder Rule [39, Theorem 4.4] then determines the right endpoint of the interval associated to M_{n_i} to be t_i for $i \geqslant 2$, leaving the right endpoint or M_{n_1} undecided. There was a choice in our ordering of the M_{n_i} of which local minimum was assigned to M_{n_1} if multiple of the M_{n_i} have $f(M_{n_i}) = f(M_{n_1})$. However, according to the Elder Rule, regardless of the choice we make, we will get the same barcode. Our inductive hypothesis is satisfied, as we have exactly one local minimum, namely, M_{n_1}, which has a right endpoint yet to be determined in the connected component of T_{t_i} containing each C_l. As before, only one interval will remain with undecided right endpoint after this procedure, and we label this right endpoint with ∞ to produce the barcode $\mathrm{PH}(T, f)$. Replacing ∞ with $\max(f)$ we can use this algorithm to compute $\mathrm{EPH}_0(T, f)$.

We can easily define the negation operation on barcodes to be the map which swaps and takes the negative value of interval endpoints. For example, we have

$$-([1,\infty) \sqcup [2,5) \sqcup [3,4)) = (-\infty,-1] \sqcup (-5,-2] \sqcup (-4,-3].$$

Similarly, we can also define a switching map S which switches the endpoints of each interval. For example,

$$S([1,\infty) \sqcup [2,5) \sqcup [3,4)) = (\infty,1] \sqcup (5,2] \sqcup (4,3].$$

The switching map does not send intervals in $\mathbb{R}$ to intervals in $\mathbb{R}$ since the right endpoint of a genuine interval will, after switching, be less than or equal to its left endpoint. Nevertheless, this is a useful notion to consider.

In the sequence of Eq. (A1), we constructed T sequentially via the sublevel sets of f. We could proceed analogously using the *superlevel sets* of f, the subgraphs T^t generated by the nodes v satisfying $f(v) \geqslant t$, this time with t decreasing with each inclusion map. The exact same procedure gives us a barcode corresponding to the sublevel sets of f, with the caveat that the so-called intervals $[x, y)$ in this barcode satisfy $x \geqslant y$. Similarly, we can make these intervals finite by replacing the $-\infty$ in the resulting barcode with $\min(f)$, and replacing the open endpoint corresponding to $-\infty$ with a closed endpoint. It is readily verified that applying persistent homology to the superlevel sets of f gives us the barcode $-S(\mathrm{PH}(T,-f))$. Applying persistent homology to the superlevel sets composed with the map $-\infty \mapsto \min(f)$, which changes the infinite endpoint to be closed, is then easily seen to be $-S(\mathrm{EPH}_0(T,-f))$.

3. Metrics on the space of persistence diagrams

We first recall the definition of a persistence diagram.

Definition 1. A **persistence diagram** is a multiset of points in $\bar{\mathbb{R}}^2$ that contains every diagonal point (x, x) for $x \in \mathbb{R}$ with infinite multiplicity. Here, $\bar{\mathbb{R}}$ denotes the extended real line $\mathbb{R} \cup \{-\infty\} \cup \{\infty\}$.

Every barcode induces a persistence diagram by the map which sends each interval with left endpoint x and right endpoint y to the point (x, y) and then includes every diagonal point (x, x) with infinite multiplicity. This map is well defined even if $y < x$ for some subcollection of intervals. We restrict our attention to persistence diagrams with finitely many off-diagonal points, each with finite multiplicity. Indeed, all finite trees T must have persistence diagrams associated to $\mathrm{TMD}(T, f)$ and $\mathrm{EPH}_0(T, f)$ of this type since T has finitely many nodes.

Given pairs of persistence diagrams D_1 and D_2, it is often useful to have a numerical measure of how different they are. Two popular such measurements of the difference between pairs of persistence diagrams are the q-Wasserstein and bottleneck distances.

Definition 2. For $q \geqslant 1$ the ***q*-Wasserstein distance** between persistence diagrams D_1 and D_2 is given by

$$W_q(D_1, D_2) := \inf_{\phi} \left(\sum_{x \in D_1} \|x - \phi(x)\|_\infty^q \right)^{1/q},$$

where the infimum is taken over all bijective maps $\phi : D_1 \to D_2$. If the term inside the infimum is never finite then we let $W_q(D_1, D_2) = \infty$. We define the **bottleneck distance** similarly by

$$d_B(D_1, D_2) = W_\infty(D_1, D_2) := \inf_{\phi} \sup_{x \in D_1} \|x - \phi(x)\|_\infty.$$

We say a bijective map $\phi : D_1 \to D_2$ satisfying

$$\sup_{x \in D_1} \|x - \phi(x)\|_\infty \leqslant \delta$$

is a **δ-matching** of D_1 and D_2.

For this definition, we take $|\infty - \infty| = 0$ and $|\infty - t| = \infty$ for any $t \in \bar{\mathbb{R}} - \{\infty\}$, and similarly define the absolute value for $-\infty$.

The bottleneck distance is shown to be an extended metric on the space of persistence diagrams with finitely many off-diagonal points in Ref. [16, p. 219]. The Wasserstein distances are shown to be extended metrics similarly.

When comparing two barcodes B_1 and B_2 we abuse notation and let $W_q(B_1, B_2)$ and $d_B(B_1, B_2)$ denote the Wasserstein and bottleneck distances of their associated persistence diagrams. It should be noted that distance these functions are not metrics on the space of barcodes, for two barcodes can have the same persistence diagram and yet differ on the openness or closedness of their endpoints. As this information is all that is lost in the mapping from barcodes to persistence diagrams, this is the only way two barcodes of bottleneck or Wasserstein distance zero can differ.

4. The TMD versus persistent homology

The TMD algorithm also produces a barcode $\mathrm{TMD}(T, f)$ from a tree T equipped with a function f [13]. We recall that the algorithm can be paraphrased as follows.

(1) Choose a branch point and identify a child branch of that branch point that attains the greatest value of f on its leaves.

(2) Detach every child branch from this branch point except for the identified branch.

(3) If there are any branch points in the resulting collection of trees, return to step 1.

(4) Label the endpoints of the resulting collection intervals with the f values associated to their endpoints.

Our goal is to connect the TMD operation to the EPH_0 operation for a certain class of functions on T including d, the intrinsic distance from the root. Prior related work, for example [40] and [14], has remarked and alluded that for functions f which increase along paths moving away from the root, the TMD and EPH_0 coincide. We provide a proof of this fact.

Proof of Theorem 1. Let l_0 be one of the leaves with the largest f value, and let $f(l_0) = L$. From the fact that f is increasing with respect to the directed edges of T, it follows that the leaves of T are the only local minima of $-f$, and so the intervals of $\mathrm{PH}(T, -f)$ bijectively correspond to leaves of T. Similarly, the TMD algorithm also associates an interval to each leaf of T. Connected components merge in the sublevel set filtration of T and $-f$ only when a branch point b is included, and the components that are merged correspond to the child branches of b. We show the theorem holds by computing $\mathrm{TMD}(T, f)$ and $\mathrm{PH}(T, -f)$ concurrently. For each iteration of the TMD algorithm, choose b, one of the branch points with the greatest value of f, in order to ensure that every detached child branch is actually an interval, and thus contains a single leaf. Then, choose one of the child branches which attains the largest value of f on its leaf l. The Elder Rule [39, Theorem 4.4] then dictates that each leaf $l' \neq l$ of a child branch of b may be associated to the interval $[-f(l'), -f(b))$ in $\mathrm{PH}(T, -f)$. Meanwhile the TMD algorithm associates to the child branch containing $l' \neq l$ the interval $(f(b), f(l')]$. The TMD algorithm then dictates that we detach the child branch containing each $l' \neq l$. This does not change the left endpoint given by the Elder Rule for an interval corresponding to any other leaf since removing the branch from b to l' does not change the minimal value of $-f$ at any of the remaining child branches of branch points of T, nor the value of $-f$ the remaining branch points of T themselves. After completing this procedure for every branch point of T, what remains to be computed is the infinite interval $[-L, \infty)$ in $\mathrm{PH}(T, -f)$ and the finite interval $[f(r), L]$ in $\mathrm{TMD}(T, f)$. From the monotonicity assumption, $f(r) = \max(-f)$, and so it is immediate that $\mathrm{TMD}(T, f) = -\mathrm{EPH}_0(T, -f)$. ■

An immediate consequence of this result is that the methods of Refs. [13,14] can record identical information for functions f satisfying the requirements of the theorem. In particular, the function d, the path distance to the root, satisfies the requirements of this theorem, and so $\mathrm{TMD}(T, d) = -\mathrm{EPH}_0(T, -d)$.

E. Topological morphology functions from persistence surfaces

For a neuron represented by a tree T, recall that the associated topological morphology function $p(t)$ returns the number of points on the neuron with an intrinsic distance of t from the root.

We can transform the persistence diagram of $\mathrm{TMD}(T, f)$ by the map $(x, y) \mapsto (x, y - x)$. If d is the intrinsic distance from the root function, we have shown in the main text that for t not an endpoint of an interval of $\mathrm{TMD}(T, d)$, $p(t)$ is the number of points in transformed persistence diagram of $\mathrm{TMD}(T, d)$ in the region given by $x < t$ and $y > t - x$, which we call R_t. Adding together two-dimensional Gaussian functions with standard deviation σ centered at each point on the transformed persistence diagram of $\mathrm{TMD}(T, d)$, each weighted by the y value of their center, we produce the persistence surface $F_\sigma(x, y)$. Theorem 2, which remains to be proven, shows that an approximation of p can be constructed from F_σ.

We will formally restate this theorem shortly, but first we will need a definition and a lemma.

Definition 3. Consider the family of boxes $B_\delta(\mu)$ of width and height δ centered around $\mu = (\mu_x, \mu_y)$. Let $g_\sigma(x, y; \mu)$ be a family of functions for positive σ satisfying the following properties:

(1) The function $g_\sigma(x, y; \mu)$ is positive, and is bounded for fixed σ.

(2) $\int_{\mathbb{R}^2} g_\sigma(x, y; \mu)\, dx\, dy = 1$.

(3) $g_\sigma(x, y; \mu) \to 0$ uniformly as $\sigma \to 0$ on $B_\delta^C(\mu)$, the complement of $B_\delta(\mu)$.

(4) $\int_{B_\delta(\mu)} g_\sigma(x, y; \mu)\, dx\, dy \to 1$ as $\sigma \to 0$.

(5) Every partial derivative of $g_\sigma(x, y; \mu)$ with respect to x is continuous, and for fixed σ and μ is bounded in absolute value both by some constant and by $M/(x^2 + y^2)$ for some other constant M.

Then we say that $g_\sigma(x, y; \mu) \in \mathcal{F}$.

Lemma 1. If $g_\sigma(x, y; \mu) \in \mathcal{F}$, then

$$\lim_{\sigma \to 0} \int_{R_t} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy \tag{A2}$$

is equal to zero if μ is in the exterior of R_t and is equal to μ_y^{-1} if μ is in the interior of R_t. Further, for each positive σ,

$$\int_{R_t} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy$$

is infinitely differentiable as a function of t.

Proof. We only use property 5 to show the differentiability condition holds at the end. From properties 2 and 4, we immediately have that $\int_{B_\delta^C(\mu)} g_\sigma(x, y; \mu)\, dx\, dy \to 0$ as $\sigma \to 0$.

We first show that the integral in the theorem converges regardless of t. For positive h let Δ_h denote the triangular subset of R_t of points (x, y) additionally satisfying that $y < h$. Let M be a bound for $g_\sigma(x, y; \mu)$ for a given σ. We have

$$\begin{aligned}\int_{\Delta_h} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy &= \int_0^h \int_{t-y}^t \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy \\ &\leqslant \int_0^h \int_{t-y}^t \frac{M}{y}\, dx\, dy = Mh,\end{aligned}$$

$$\int_{R_t - \Delta_h} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy \leqslant \int_{R_t - \Delta_h} \frac{g_\sigma(x, y; \mu)}{h}\, dx\, dy \leqslant \frac{1}{h}.$$

To establish the limiting value in the theorem, first consider $g_\sigma(x, y; \mu)$ fixing μ in the exterior of R_t. For such μ, there

exists a δ sufficiently small that $B_\delta(\mu)$ is disjoint from R_t. Let $S(\sigma)$ be the supremum of $g_\sigma(x, y; \mu)$ on $B_\delta^C(\mu)$ as a function of σ. Property 3 implies that this approaches 0 as σ does. We have

$$\begin{aligned}\int_{\Delta_h} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy &= \int_0^h \int_{t-y}^t \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy \\ &\leqslant \int_0^h \int_{t-y}^t \frac{S(\sigma)}{y}\, dx\, dy = hS(\sigma),\end{aligned}$$

$$\int_{R_t-\Delta_h} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy \leqslant \int_{R_t-\Delta_h} \frac{g_\sigma(x, y; \mu)}{h}\, dx\, dy \leqslant \frac{1}{h}.$$

Letting $h = S(\sigma)^{-1/2}$, we observe

$$0 \leqslant \int_{R_t} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy \leqslant 2S(\sigma)^{1/2},$$

which tends to zero as σ does.

If on the other hand μ lies in R_t, again we can choose δ small enough that $B_\delta(\mu)$ is a subset of R_t. Define $S(\sigma)$ as before. Once again we obtain bounds

$$\begin{aligned}\int_{\Delta_h - B_\delta(\mu)} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy &\leqslant \int_{\Delta_h} \frac{S(\sigma)}{y}\, dx\, dy \\ &= \int_0^h \int_{t-y}^t \frac{S(\sigma)}{y}\, dx\, dy \\ &= hS(\sigma),\end{aligned}$$

$$\begin{aligned}\int_{R_t-\Delta_h\cup B_\delta(\mu)} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy &\leqslant \int_{R_t-\Delta_h} \frac{g_\sigma(x, y; \mu)}{h}\, dx\, dy \\ &\leqslant \frac{1}{h}.\end{aligned}$$

Combining these two results and letting $h = S(\sigma)^{-1/2}$, we observe

$$0 \leqslant \int_{R_t - B_\delta(\mu)} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy \leqslant 2S(\sigma)^{1/2},$$

which approaches zero as σ approaches zero. Meanwhile, we also have

$$\begin{aligned}\int_{B_\delta(\mu)} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy &\leqslant \int_{B_\delta(\mu)} \frac{g_\sigma(x, y; \mu)}{\mu_y - \delta}\, dx\, dy \\ &\leqslant \frac{1}{\mu_y - \delta},\end{aligned}$$

$$\begin{aligned}\int_{B_\delta(\mu)} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy &\geqslant \int_{B_\delta(\mu)} \frac{g_\sigma(x, y; \mu)}{\mu_y + \delta}\, dx\, dy \\ &\to \frac{1}{\mu_y + \delta},\end{aligned}$$

with the limit in the last line taken as $\sigma \to 0$. Since δ can be taken arbitrarily small, this implies

$$\int_{B_\delta(\mu)} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy \to \frac{1}{\mu_y}.$$

Hence,

$$\int_{R_t} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy \to \frac{1}{\mu_y}.$$

In summary, we have shown that the above integral approaches μ_y^{-1} when μ is interior to R_t and approaches zero when μ is exterior to R_t.

All that remains to be shown is the differentiability statement of the lemma, i.e., we must show that the integral

$$I(t; \mu) := \int_{R_t} \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy = \int_0^\infty \int_{t-y}^t \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy$$

is an infinitely differentiable function of t.

To begin, notice that by continuity of g we can extend the integrand to $y = 0$ via

$$\lim_{y\to 0} \int_{t-y}^t \frac{g_\sigma(x, y; \mu)}{y}\, dx = g_\sigma(t, y; \mu).$$

Further, we have the estimate for positive C:

$$\begin{aligned}\int_C^\infty \int_{t-y}^t \frac{g_\sigma(x, y; \mu)}{y}\, dx\, dy &\leqslant \int_C^\infty \int_{t-y}^t \frac{g_\sigma(x, y; \mu)}{C}\, dx\, dy \\ &\leqslant \int_{\mathbb{R}^2} \frac{g_\sigma(x, y; \mu)}{C}\, dx\, dy \leqslant \frac{1}{C},\end{aligned}$$

showing uniform convergence of the same integral with C taken to be 0, since g is positive valued. We also have that

$$\begin{aligned}&\frac{\partial^n}{\partial t^n} \int_{t-y}^t \frac{g_\sigma(x, y; \mu)}{y}\, dx \\ &\quad= \frac{\partial^{n-1}}{\partial t^{n-1}} \frac{g_\sigma(t, y; \mu) - g_\sigma(t - y, y; \mu)}{y} \\ &\quad= \frac{1}{y}\left[\frac{\partial^{n-1}}{\partial t^{n-1}} g_\sigma(t, y; \mu) - \frac{\partial^{n-1}}{\partial t^{n-1}} g_\sigma(t - y, y; \mu)\right].\end{aligned}$$

This expression is bounded by any constant bounding of $|\frac{\partial^n}{\partial t^n} g_\sigma(t, y; \mu)|$. Hence, applying $\int_0^\infty dy$ to this expression, the lower limit converges uniformly. Meanwhile, letting M be a bound for the integral of $|\frac{\partial^{n-1}}{\partial t^{n-1}} g_\sigma(t, y; \mu)|$ over linear domains in t and y, we observe that the upper limit converges uniformly as well since

$$\begin{aligned}&\left|\int_C^\infty \frac{1}{y}\left[\frac{\partial^{n-1}}{\partial t^{n-1}} g_\sigma(t, y; \mu) - \frac{\partial^{n-1}}{\partial t^{n-1}} g_\sigma(t - y, y; \mu)\right] dy\right| \\ &\quad\leqslant \int_C^\infty \frac{1}{y}\left[\left|\frac{\partial^{n-1}}{\partial t^{n-1}} g_\sigma(t, y; \mu)\right| + \left|\frac{\partial^{n-1}}{\partial t^{n-1}} g_\sigma(t - y, y; \mu)\right|\right] dy \\ &\quad\leqslant \int_C^\infty \frac{1}{y}\left[\frac{M}{t^2 + y^2} + \frac{M}{(t - y)^2 + y^2}\right] dy \\ &\quad\leqslant \int_C^\infty \frac{2M}{y^3}\, dy \leqslant \frac{4M}{C^2}.\end{aligned}$$

From this we know we can interchange the partial derivative and integral in the expression for $I(t; \mu)$, giving us the formula

$$\begin{aligned}\frac{\partial^n}{\partial t^n} I(t; \mu) = \int_0^\infty \frac{1}{y}\Bigg[&\frac{\partial^{n-1}}{\partial t^{n-1}} g_\sigma(t, y; \mu) \\ &- \frac{\partial^{n-1}}{\partial t^{n-1}} g_\sigma(t - y, y; \mu)\Bigg] dy,\end{aligned}$$

completing the proof. ■

With this Lemma we can easily show Theorem 2.

Theorem 2. Let p be the topological morphology function associated to a neuron represented by a tree T. Let $d : N(T) \to [0, \infty)$ be the intrinsic distance to the soma. Let $F_\sigma(x, y)$ be the persistence surface corresponding to TMD(T, d) constructed with Gaussian functions of standard deviation σ. For any generic positive number t,

$$p(t) = \lim_{\sigma \to 0} \int_{R_t} \frac{F_\sigma(x, y)}{y}\, dx\, dy. \tag{3}$$

Further the integral in the above expression converges and is an infinitely differentiable function of t for all $\sigma > 0$.

Proof of Theorem 2. Let $\mu_1, \cdots, \mu_N$ be the coordinates of the off-diagonal points in the transformed persistence diagram of TMD(T, f). The persistence surface is then given by the formula

$$F_\sigma(x, y) = \sum_{i=1}^{N} (\mu_i)_y\, g_\sigma(x, y; \mu_i). \tag{A3}$$

Each function $g_\sigma(x, y; \mu_i)$ is a two-dimensional Gaussian function, for which the requirements of Lemma 1 are elementary properties. Each statement of theorem 2 is immediate from the fact that F_σ is a finite linear combination of functions satisfying the requirements of Lemma 1. Indeed, in light of Eq. (A3), Eq. (3) is clearly infinitely differentiable from the previous lemma. When t is generic with respect to B, it is easily seen that each μ_i is not on the boundary of R_t. Applying Lemma 1 and Eq. (A3), we see that the limit

$$\lim_{\sigma \to 0} \int_{R_t} \frac{F_\sigma(x, y)}{y}\, dx\, dy$$

evaluates to the number of μ_i in R_t, which is equal to $p(t)$. ■

Let T be a tree with an embedding P to $\mathbb{R}^n$ and f be the Euclidean distance of each point represented by v to the root r. Recall that we define the Sholl function $s(t)$ to be the number times the n-sphere of radius t centered about $P(r)$ intersects with $P(T)$. Let B be the barcode

$$\mathrm{EPH}_0(T, f) \sqcup -S(\mathrm{EPH}_0(T, -f)), \tag{A4}$$

the disjoint union of barcodes given by persistent homology of the superlevel and sublevel sets of f. We say that t is generic if it is not an endpoint of an interval in B. Let D_B be the persistence diagram associated to B. If instead of assuming the embedding of edges of T in $\mathbb{R}^n$ is linear, we assume that P is such that f is weakly increasing or decreasing on edges, Li *et al.* show in Ref. [14, Sec. 2.4] that the value $s(t)$ can be recovered from D_B for any generic t. For the remainder of this section, we only consider T with such an embedding. For generic t between 0 and $\max(f)$, Li *et al.* prove that the Sholl function $s(t)$ is the number of points above and to the left or below and to the right of (t, t) in D_B minus one. More formally written, if Q_t is the set of (x, y) such that ($x < t$ and $y > t$) or ($x > t$ and $y < t$), then

$$s(t) = \left|B \cap Q_t\right| - 1. \tag{A5}$$

If t is generic, but t is not between 0 and $\max(f)$ it is easily seen that the left side of this equation is 0 while the right side of this equation is -1. We show how $s(t)$ may be calculated from D_B visually in Fig. 12.

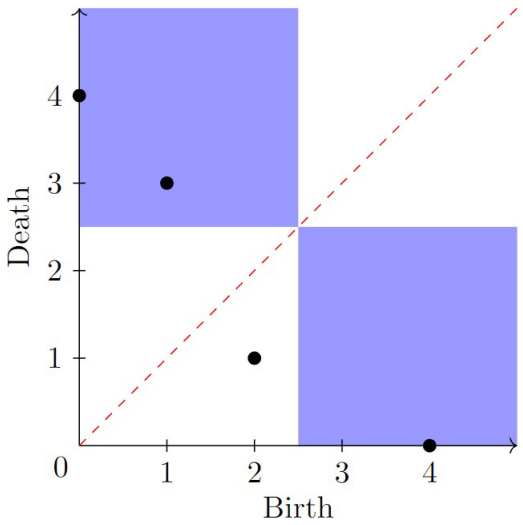

FIG. 12. Calculating the $s(2.5)$ from the persistence diagram D_T. Shown in blue is the region $Q_{2.5}$, within which there are three points. Hence, $s(2.5) = 2$.

One may define a persistence surface F_σ for the barcode B defined in Eq. (A4) by adding a two-dimensional Gaussian function for each point in D_B, with multiplicity, each weighted by the vertical distance of the point from the diagonal.[3] We can use our Lemma 1 to show that an approximate Sholl function can be recovered from this persistence surface.

Theorem 3. Let T be a rooted tree with Sholl function s and f be the associated Euclidean distance from the root function. Assume that between adjacent vertices, the embedding of T in $\mathbb{R}^2$ or $\mathbb{R}^3$ is such that f is either weakly increasing or decreasing along each edge. Let B be the barcode given by Eq. (A4) and D_B be the associated persistence diagram. Let $F_\sigma(x, y)$ be the persistence surface corresponding to D_B constructed with Gaussian functions of standard deviation σ, without the transformation step $(x, y) \mapsto (x, y - x)$. For any generic positive number t between 0 and $\max(f)$,

$$s(t) = \lim_{\sigma \to 0} \int_{Q_t} \frac{F_\sigma(x, y)}{|x - y|}\, dx\, dy - 1.$$

Otherwise, if t is generic, the left side of this equation is 0 and the right side of this expression is -1. Further, the integral in the above expression converges and is an infinitely differentiable function of t for all $\sigma > 0$.

Proof. Let $\mu_1, \cdots, \mu_N$ be the coordinates of the off-diagonal points in D_B, with x and y values $(\mu_i)_x$ and $(\mu_i)_y$ respectively. By definition the persistence surface F_σ is given by a weighted sum of Gaussians:

$$F_\sigma(x, y) = \sum_{i=1}^{N} |(\mu_i)_y - (\mu_i)_x|\, g_\sigma(x, y; \mu_i).$$

We denote by Q_t^+ the portion of Q_t satisfying $y > x$ and similarly denote by Q_t^- the portion of Q_t satisfying $x > y$. We

[3] It is common to skip the transformation step when there are points below the diagonal in the initial barcode.

have the equation

$$\int_{Q_t} \frac{F_\sigma(x,y)}{|x-y|}\,dx\,dy = \int_{Q_t^+} \frac{F_\sigma(x,y)}{y-x}\,dx\,dy + \int_{Q_t^-} \frac{F_\sigma(x,y)}{x-y}\,dx\,dy.$$

Examining the first term, we have

$$\begin{aligned}\int_{Q_t^+} \frac{F_\sigma(x,y)}{y-x}\,dx\,dy &= \int_0^\infty \int_{-\infty}^t \frac{F_\sigma(x,y)}{y-x}\,dx\,dy \\ &= \int_0^\infty \int_{t-v}^t \frac{F_\sigma(u,u+v)}{v}\,du\,dv\end{aligned}$$

via the transform $(x,y) \mapsto (x, y-x) := (u,v)$. Noting that linear transformations of Gaussian density functions are still Gaussian density functions and that this last integral is over the region R_t in the uv plane, we have

$$\begin{aligned}&\int_0^\infty \int_{t-v}^t \frac{F_\sigma(u,u+v)}{v}\,du\,dv \\ &\quad= \sum_{i=1}^N |(\mu_i)_y - (\mu_i)_x| \int_{R_t} \frac{g_\sigma(u,u+v;\mu_i)}{v}\,du\,dv.\end{aligned}$$

We may now apply Lemma 1 as each g_σ is Gaussian with mean $((\mu_i)_x, (\mu_i)_y - (\mu_i)_x)$ in the uv plane and therefore upholds the requirements of the lemma. Thus the integrals on the right are well defined and infinitely differentiable. Consider the ith of these integrals on the right. Taking the limit $\sigma \to 0$ this integral evaluates to $((\mu_i)_y - (\mu_i)_x)$ if $((\mu_i)_x, (\mu_i)_y - (\mu_i)_x)$ is interior to R_t and 0 if it is exterior to R_t. Equivalently, the integral evaluates to $((\mu_i)_y - (\mu_i)_x)$ if μ_i is interior to Q_t^+ and 0 if it is exterior to Q_t^+.

No μ_i lies on the boundary of Q_t^+. Hence, the limit

$$\lim_{\sigma\to 0} \int_{Q_t^+} \frac{F_\sigma(x,y)}{|x-y|}\,dx\,dy$$

evaluates to the number of μ_i in Q_t^+ for generic t.

Similarly, we may show via the transformation $(x,y) \mapsto (y, x-y)$ that the limit

$$\lim_{\sigma\to 0} \int_{Q_t^-} \frac{F_\sigma(x,y)}{|x-y|}\,dx\,dy$$

is the number of μ_i in Q_t^- for generic t, and that the integral here is always well defined and infinitely differentiable regardless of t.

Hence the integral

$$\int_{Q_t} \frac{F_\sigma(x,y)}{|x-y|}\,dx\,dy = \int_{Q_t^+} \frac{F_\sigma(x,y)}{|x-y|}\,dx\,dy + \int_{Q_t^-} \frac{F_\sigma(x,y)}{|x-y|}\,dx\,dy$$

is well defined and infinitely differentiable for any σ.

Now consider generic t. We have

$$\begin{aligned}\lim_{\sigma\to 0} \int_{Q_t} \frac{F_\sigma(x,y)}{|x-y|}\,dx\,dy - 1 &= \lim_{\sigma\to 0} \int_{Q_t^+} \frac{F_\sigma(x,y)}{|x-y|}\,dx\,dy \\ &\quad + \lim_{\sigma\to 0} \int_{Q_t^-} \frac{F_\sigma(x,y)}{|x-y|}\,dx\,dy - 1,\end{aligned}$$

which is the number of points in the barcode that also lie in Q_t minus one. This value has been shown to be $s(t)$ in Ref. [14] when t is between 0 and $\max(f)$. When t is not between 0 and $\max(f)$, there can be no points in Q_t^+ or Q_t^-, and so this expression evaluates to -1. Clearly for such t, $s(t) = 0$. ■

[1] M. London and M. Häusser, Dendritic computation, Annu. Rev. Neurosci. **28**, 503 (2005).

[2] W. Rall, R. Burke, T. Smith, P. G. Nelson, and K. Frank, Dendritic location of synapses and possible mechanisms for the monosynaptic epsp in motoneurons., J. Neurophysiol. **30**, 1169 (1967).

[3] E. E. Fetz and B. Gustafsson, Relation between shapes of post-synaptic potentials and changes in firing probability of cat motoneurones, J. Physiol. **341**, 387 (1983).

[4] W. B. Grueber, L. Y. Jan, and Y. N. Jan, Tiling of the drosophila epidermis by multidendritic sensory neurons, Development **129**, 2867 (2002).

[5] G. A. Ascoli, L. Alonso-Nanclares, S. A. Anderson, G. Barrionuevo, R. Benavides-Piccione, A. Burkhalter, G. Buzsáki, B. Cauli, J. DeFelipe, A. Fairén, D. Feldmeyer, G. Fishell, Y. Fregnac, T. F. Freund, D. Gardner, E. P. Gardner, J. H. Goldberg, M. Helmstaedter, S. Hestrin, F. Karube, Z. F. Kisvárday, B. Lambolez, D. A. Lewis *et al.*, Petilla terminology: nomenclature of features of gabaergic interneurons of the cerebral cortex, Nat. Rev. Neurosci. **9**, 557 (2008).

[6] J. DeFelipe, P. L. López-Cruz, R. Benavides-Piccione, C. Bielza, P. Larrañaga, S. Anderson, A. Burkhalter, B. Cauli, A. Fairén, D. Feldmeyer *et al.*, New insights into the classification and nomenclature of cortical gabaergic interneurons, Nat. Rev. Neurosci. **14**, 202 (2013).

[7] S. Laturnus, D. Kobak, and P. Berens, A systematic evaluation of interneuron morphology representations for cell type discrimination, Neuroinformatics **18**, 591 (2020).

[8] R. Scorcioni, S. Polavaram, and G. A. Ascoli, L-measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies, Nat. Protoc. **3**, 866 (2008).

[9] G. S. Jefferis, C. J. Potter, A. M. Chan, E. C. Marin, T. Rohlfing, C. R. Maurer Jr, and L. Luo, Comprehensive maps of drosophila higher olfactory centers: spatially segregated fruit and pheromone representation, Cell **128**, 1187 (2007).

[10] J. Snider, A. Pillai, and C. F. Stevens, A universal property of axonal and dendritic arbors, Neuron **66**, 45 (2010).

[11] C. M. Teeter and C. F. Stevens, A general principle of neural arbor branch density, Curr. Biol. **21**, 2105 (2011).

[12] D. A. Sholl, Dendritic organization in the neurons of the visual and motor cortices of the cat, J. Anatomy **87**, 387 (1953).

[13] L. Kanari, P. Dłotko, M. Scolamiero, R. Levi, J. Shillcock, K. Hess, and H. Markram, A topological representation of branching neuronal morphologies, Neuroinformatics **16**, 3 (2018).

[14] Y. Li, D. Wang, G. A. Ascoli, P. Mitra, and Y. Wang, Metrics for comparing neuronal tree shapes based on persistent homology, PLoS ONE **12**, e0182184 (2017).

[15] G. Carlsson, Topology and data, Bull. Am. Math. Soc. **46**, 255 (2009).

[16] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction* (American Mathematical Society, Providence, Rhode Island, 2010).

[17] R. Ghrist, Barcodes: The persistent topology of data, Bull. Am. Math. Soc. **45**, 61 (2008).

[18] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, Persistence images: A stable vector representation of persistent homology, J. Mach. Learn. Res. **18** (2017).

[19] L. Kanari, S. Ramaswamy, Y. Shi, S. Morand, J. Meystre, R. Perin, M. Abdellah, Y. Wang, K. Hess, and H. Markram, Objective morphological classification of neocortical pyramidal cells, Cereb. Cortex **29**, 1719 (2019).

[20] Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue, and Y. Nishiura, Hierarchical structures of amorphous solids characterized by persistent homology, Proc. Natl. Acad. Sci. USA **113**, 7035 (2016).

[21] D. Beers, H. A. Harrington, and A. Goriely, Stability of topological descriptors for neuronal morphology, arXiv:2211.09058.

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

[23] K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer, Fréchet means for distributions of persistence diagrams, Discr. Computat. Geom. **52**, 44 (2014).

[24] E. Munch, K. Turner, P. Bendich, S. Mukherjee, J. Mattingly, and J. Harer, Probabilistic fréchet means for time varying persistence diagrams, arXiv:1307.6530.

[25] M. J. Sulkowski, S. C. Iyer, M. S. Kurosawa, E. P. R. Iyer, and D. N. Cox, Turtle functions downstream of cut in differentially regulating class specific dendrite morphogenesis in drosophila, PLoS ONE **6**, e22611 (2011).

[26] G. A. Ascoli, Mobilizing the base of neuroscience data: the case of neuronal morphologies, Nat. Rev. Neurosci. **7**, 318 (2006).

[27] R. Khalil, S. Kallel, A. Farhat, and P. Dlotko, Topological sholl descriptors for neuronal clustering and classification, PLoS Comput. Biol. **18**, e1010229 (2022).

[28] M. K. Bondulich, T. Guo, C. Meehan, J. Manion, T. Rodriguez Martin, J. C. Mitchell, T. Hortobagyi, N. Yankova, V. Stygelbout, J.-P. Brion *et al.*, Tauopathy induced by low level expression of a human brain-derived tau fragment in mice is rescued by phenylbutyrate, Brain **139**, 2290 (2016).

[29] D. Goniotaki, F. Tamagnini, L. Biasetti, S.-L. Rumpf, C. Troakes, S. J. Pollack, S. Ukwesa, H. Sun, L. C. Serpell, W. Noble, K. Staras, and D. P. Hanger, Tau-mediated synaptic dysfunction is coupled with HCN channelopathy, bioRxiv, 2020.11.08.369488 (2023).

[30] J. L. Crimins, A. B. Rocher, and J. I. Luebke, Electrophysiological changes precede morphological changes to frontal cortical pyramidal neurons in the rtg4510 mouse model of progressive tauopathy, Acta Neuropathol. **124**, 777 (2012).

[31] D. L. Dickstein, H. Brautigam, S. D. Stockton, J. Schmeidler, and P. R. Hof, Changes in dendritic complexity and spine morphology in transgenic mice expressing human wild-type tau, Brain Struct. Function **214**, 161 (2010).

[32] G. Einstein, R. Buranosky, and B. J. Crain, Dendritic pathology of granule cells in alzheimer's disease is unrelated to neuritic plaques, J. Neurosci. **14**, 5077 (1994).

[33] E. Falke, J. Nissanov, T. W. Mitchell, D. A. Bennett, J. Q. Trojanowski, and S. E. Arnold, Subicular dendritic arborization in alzheimer's disease correlates with neurofibrillary tangle density, Am. J. Pathol. **163**, 1615 (2003).

[34] N. Golovyashkina III, L. Penazzi, C. Ballatore, A. B. Smith, L. Bakota, and R. Brandt, Region-specific dendritic simplification induced by aβ, mediated by tau via dysregulation of microtubule dynamics: a mechanistic distinct event from other neurodegenerative processes, Mol. Neurodegen. **10**, 60 (2015).

[35] W. C. Risher, T. Ustunkaya, J. Singh Alvarado, and C. Eroglu, Rapid golgi analysis method for efficient and unbiased classification of dendritic spines, PLoS ONE **9**, e107591 (2014).

[36] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, A roadmap for the computation of persistent homology, EPJ Data Science **6**, 17 (2017).

[37] A. Zomorodian and G. Carlsson, Computing persistent homology, Discr. Computat. Geom. **33**, 249 (2005).

[38] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, Extending persistence using poincaré and lefschetz duality, Found. Computat. Math. **9**, 79 (2009).

[39] C. Cai, W. Kim, F. Mémoli, and Y. Wang, Elder-rule-staircodes for augmented metric spaces, SIAM J. Appl. Alg. Geom. **5**, 417 (2021).

[40] L. Kanari, A. Garin, and K. Hess, From trees to barcodes and back again: theoretical and statistical perspectives, Algorithms **13**, 335 (2020).