# Distance-from-flats persistent homology transform:

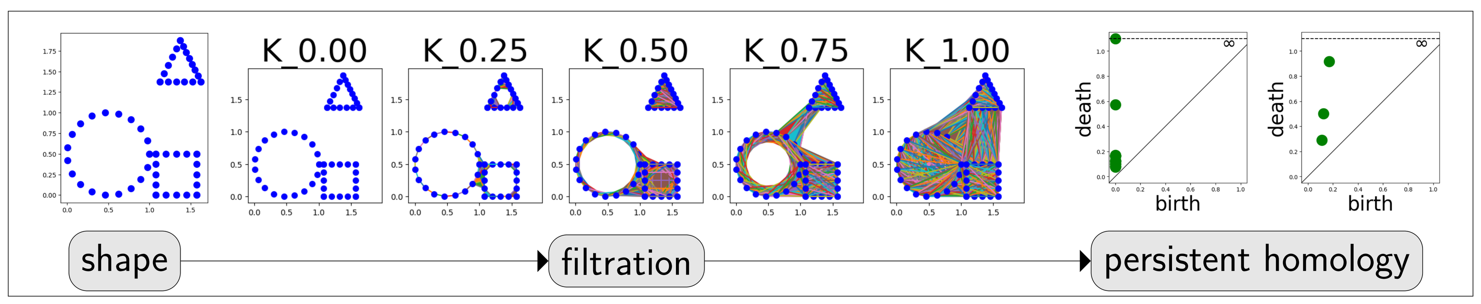
Adam Onus, Nina Otter, Renata Turkeš

## Shape analysis

Shape classification is a difficult problem that plays a crucial role in understanding and recognising physical structures and objects, image processing, and computer vision. Since topology is the branch of mathematics that studies shape, it has inspired the new field of study called topological data analysis (TDA), which, as the name suggests, aims to analyse data by studying its shape. The main tool in TDA, persistent homology (PH), captures topological and geometric features that are richer than the scalar-valued quantities that are often used in shape analysis and which struggle to describe the shape sufficiently well.

### What is persistent homology?

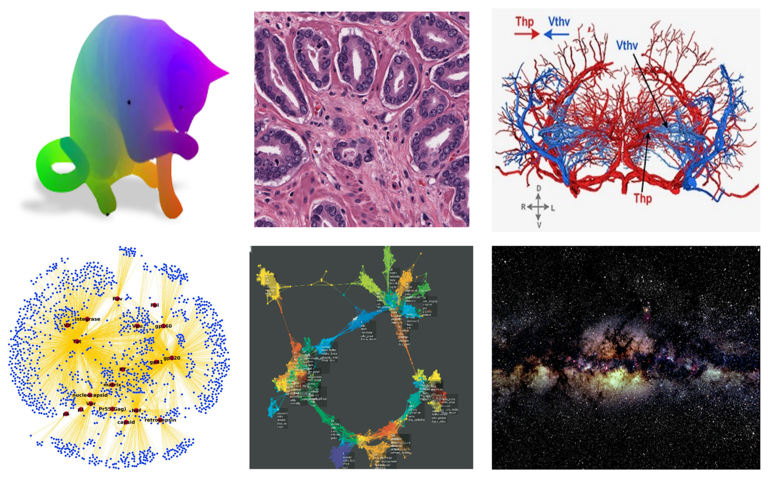
Persistent homology (PH) captures information about $k$-dimensional cycles: connected components, loops, voids and higher-dimensional voids. More precisely, PH tracks how these homological structures persist in a so-called filtration $\{X_r\}_{r\in\mathbb{R}}$, a nested family of spaces that approximate the shape $X$ at any given scale $r$. It is commonly summarized with a persistence diagram (PD), a scatter plot of "birth" and "death" values of each cycle within the filtration.




For the example shape, PD in homology degree $0$ and $1$ respectively sees the two connected components and the three loops, and the birth and death times capture additional geometric information such as the size of the loops.

### ... and why is it cool?

Persistent homology lies at the crossroads of algebraic topology and computational geometry, offering insights beyond traditional statistical or machine learning techniques, with applications across many disciplines, such as medicine (to detect cancer or other anomalies), biology (to understand protein structure or brain activity), materials science (to study the porousness of materials), or astronomy (to analyze the large-scale structure of the universe).



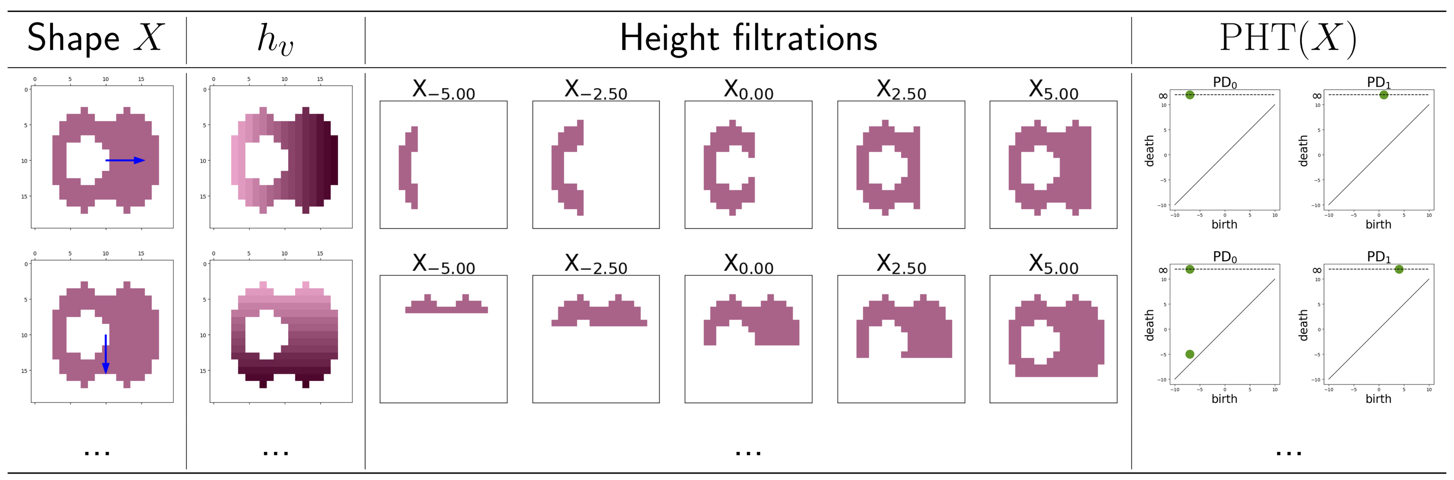
In applications, persistent homology is commonly a tool for:

- feature extraction, or
- data exploration.

### What is the classical persistent homology transform?

For a shape $X$, the classical persistent homology transform, $\mathrm{PHT}(X)$ [1], consists of persistence diagrams with respect to the height filtration function $h_v(x) = x \cdot v$ for *any* direction $v$, and *any* homological degree $k$:

$$\begin{aligned}\mathrm{PHT}(X)\colon \mathbb{S}^{n-1} &\to \mathcal{D}^n\\ v \quad &\mapsto \Big(\mathrm{PD}_0(X), \mathrm{PD}_1(X), \ldots, \mathrm{PD}_{n-1}(X)\Big),\end{aligned}$$



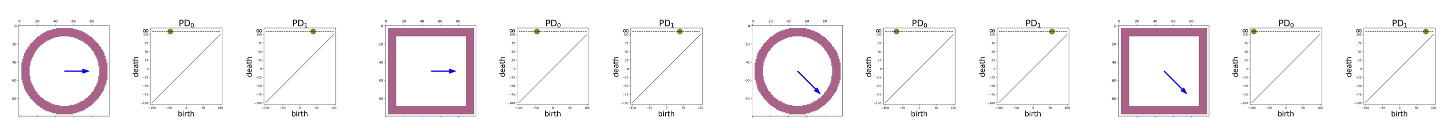

### ... and why is it cooler?

As persistent homology captures information about $k$-dimensional cycles, it provides *some* insights into the topology and geometry, but some information is lost. For instance, PH might not be able to discriminate between a circle, square or a triangle. On the other hand, the classical PHT, which in generalised notation we denote $\mathrm{PHT}_{\mathbb{S}^{n-1},h}$, has been shown to completely describe a shape:

**Theorem 1: $\mathrm{PHT}_{\mathbb{S}^{n-1},h}$ is injective [2, 3]**

Let $CS(\mathbb{R}^n)$ be the family of constructible sets in $\mathbb{R}^n$. The persistent homology transform $\mathrm{PHT}_{\mathbb{S}^{n-1},h} : CS(\mathbb{R}^n) \to C(\mathbb{S}^{n-1}, \mathcal{D}^n)$ is injective.

In other words, the classical height PHT is a sufficient shape statistic. For instance, the circle and the square, or the MNIST images of handwritten digits "0", "6" and "9", can have the same PH with respect to the greyscale, or height filtration "from left to right", but considering multiple filtrations or "directions" with PHT can help to differentiate between the shapes.



### What is the generalised persistent homology transform?

**Definition 1: Generalised persistent homology transform $\mathrm{PHT}_{\mathbb{P},f}$**

Let $X \subset \mathbb{R}^n$ be a constructible set. Let further $\mathbb{P}$ be a topological space, and $\{f_P\}_{P\in\mathbb{P}}$ be a family of functions $f_P : X \to \mathbb{R}$. The persistent homology transform of $X$ with parameter space $\mathbb{P}$ and filtration functions $\{f_P\}_{P\in\mathbb{P}}$ is the function

$$\begin{aligned}\mathrm{PHT}_{\mathbb{P},f}(X)\colon \mathbb{P} &\to \mathcal{D}^n\\ P &\mapsto (\mathrm{PD}_0(X, f_P), \mathrm{PD}_1(X, f_P), \ldots, \mathrm{PD}_{n-1}(X, f_P)),\end{aligned}$$

where $\mathrm{PD}_k(X, f_P)$ is the persistence diagram of $X$ in homology degree $k$, with respect to the sublevel-set filtration of the function $f_P$.

We focus on the following special cases of PHTs:

**Definition 2: Distance-from-flats persistent homology transform $\mathrm{PHT}_{\mathbb{AG}(m,n),\mathrm{d}}$**

The distance-from-flats persistent homology transform $\mathrm{PHT}_{\mathbb{AG}(m,n),\mathrm{d}}$ is the PHT where the domain is the affine Grassmannian space, $\mathbb{P} = \mathbb{AG}(m, n)$, and the filtration functions $f_P(x) = \mathrm{d}(x, P)$ encode the distance from $m$-flats $P \in \mathbb{AG}(m, n)$. We zoom into PHT with respect to the distance from hyperplanes ($m = n - 1$), lines ($m = 1$) and points ($m = 0$), which we refer to as the *slabbed, tubular* and *radial persistent homology transform*, respectively.

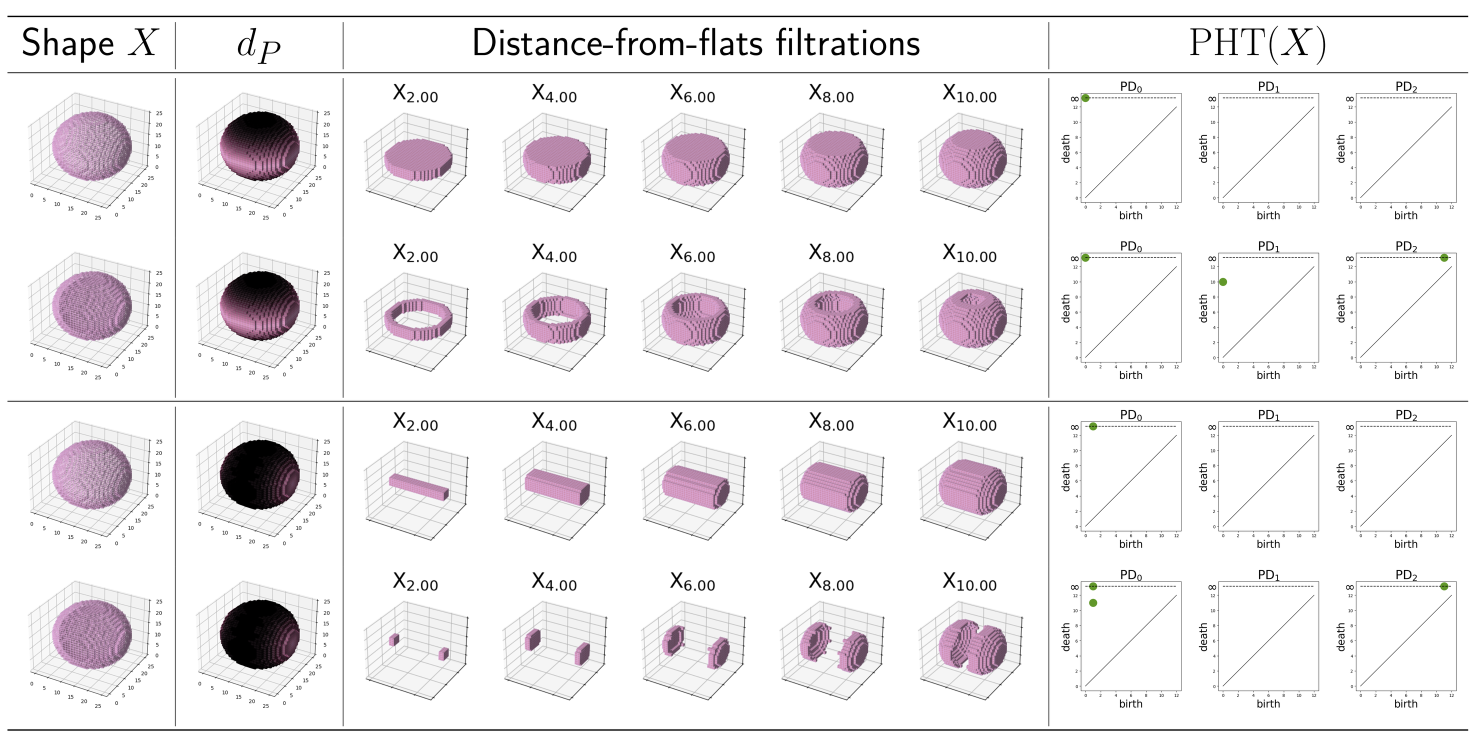
### ... and why is it the coolest?

The generalised persistent homology transform $\mathrm{PHT}_{\mathbb{P},f}$ provides a broader framework for probing shapes, that opens doors to new research on PHT with respect to some other interesting families of filtration functions, that go beyond the classical height function. Most importantly, for the distance-from-flats persistent homology transform that is the focus of this work, we show:

**Theorem 2: Truncated $\mathrm{PHT}_{\mathbb{AG}(m,n),\mathrm{d}}$ is injective**

Let $CS(\mathbb{R}^n)$ be the family of constructible sets in $\mathbb{R}^n$, and $m \geq 1$. Distance-from-flats persistent homology transform $\mathrm{PHT}_{\mathbb{AG}(m,n),\mathrm{d}} : CS(\mathbb{R}^n) \to C(\mathbb{AG}(m, n), \mathcal{D}^n)$, *truncated to homological dimensions* $k \in \{0, 1, \ldots, m - 1\}$, is injective.

In particular, the tubular $\mathrm{PHT}_{\mathbb{AG}(1,n),\mathrm{d}}$, which "shoves tubes through shapes" completely describes the shape, even when truncated to homology degree $0$. This comes with significant computational advantage since it easy to compute PH in degree $0$ in near-linear time with respect to the number $N$ of simplices, whereas the standard algorithm is $\mathcal{O}(N^3)$. Therefore, the tubular $\mathrm{PHT}_{\mathbb{AG}(1,n),\mathrm{d}}$ is an efficient and sufficient shape statistic.




More broadly, for $m' < m$, the distance-from-flats PHT on $\mathbb{AG}(m', n)$ is more efficient that PHT on $\mathbb{AG}(m, n)$. For instance, $\mathrm{PD}_0(X)$ with respect to the slabbed filtration functions, i.e., distance from planes (top rows) cannot discriminate between a ball- and sphere-like shape: there is always the one connected components in the filtration for both shapes; higher homological dimensions are needed. However, $\mathrm{PD}_0(X)$ with respect to the tubular filtration functions, i.e., distance from lines (bottom rows) is sufficient to differentiate the two shapes: the sphere sees a second connected component in the filtration.

## Open questions

- Which (size of) *finite* subset of $\mathbb{AG}(m, n)$ leads to an injective transform?
- What is the computational trade-off between calculating PH up to degree $m - 1$, compared to sampling flats from a higher-dimensional parameter space $\mathbb{AG}(m, n)$?
- How to deal with the instability of PHT?
- How meaningful is it to extend to distance from non-linear subspaces of Euclidean space?
- What are the properties of PHT on non-Euclidean shapes, such as weighted networks or subsets of hyperbolic space?

## References

[1] Katharine Turner, Sayan Mukherjee, and Doug M. Boyer, *Persistent homology transform for modeling shapes and surfaces*, Information and Inference: A Journal of the IMA (2014).

[2] Justin Curry, Sayan Mukherjee, and Katharine Turner, *How many directions determine a shape and other sufficiency results for two topological transforms*, Transactions of the American Mathematical Society, Series B (2022).

[3] Robert Ghrist, Rachel Levanger, and Huy Mai, *Persistent homology and Euler integral transforms*, Journal of Applied and Computational Topology (2018).

Curious to learn more?
Tap into the full paper!

Scan Me!