

# Persistent Homology for Data of Plant Morphology

Yuqing Xing

Adviser: Yifei Zhu

Adviser: Zhen Huan

May 24, 2021

# Outline

- 1 Motivation
- 2 Introduction
- 3 Application to Plant Morphology Analysis
- 4 References

# Outline

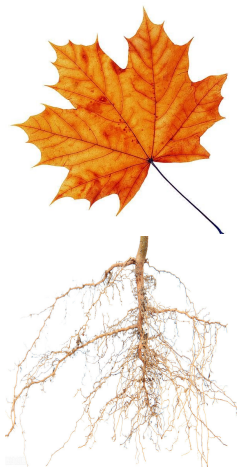
- 1 Motivation
- 2 Introduction
- 3 Application to Plant Morphology Analysis
- 4 References

# Motivation

The local morphology of plants, such as roots, branches and leaves, has obvious geometric and topological features.

Most of the leaves are flat, hence we only need to study the 2D shape of leaves. For roots, capturing its morphology can be difficult sometimes, because roots are highly branched structure lacking easily identifiable nodes and rarely fully quantified. But by using topological methods(e.g., PH, TDA), we can distinguish the root structure of plants.

# Motivation



# Motivation

Mao Li presented a morphometric technique based on topology, using a persistent homology framework, to measure the morphology of leaves and classify them by plant family(2017). For a given topological feature, PH measures the persistence of components of the feature across the scales of a filtration set, then creates a persistence barcode that can be used to compare the overall topological similarity between objects.

Combined with analysis methods (e.g., canonical variant analysis (CVA), principal component analysis(PCA)), persistent homology can be used to classify the morphology of plants. Topological features provide a framework to quantify leaflets, roots and branching architectures of plants comprehensively.

# Outline

- 1 Motivation
- 2 Introduction
- 3 Application to Plant Morphology Analysis
- 4 References

# Persistent Homology

**A Filtration** is a collection  $K = \{K_\lambda\}_{\lambda \geq 0}$  of simplicial complexes with  $K_\lambda \subset K_{\lambda'}$  for all  $\lambda \leq \lambda'$ .

Let  $(X, d)$  be a metrix space of finite points, define the **Rips complex**:

$$R_\lambda(X, d) = \{(x_0, x_1, \dots, x_n) \in X \mid d(x_i, x_j) < \lambda, 0 \leq i, j \leq n, n \in N\}$$

**A Rips filtration** is an increasing sequence of sets with the form  $R(X, d) = \{R_{\lambda_k}(X, d) \mid k = 1, \dots, N, 0 < \lambda_k < \lambda_{k+1}\}$



# Persistent Homology

## Two ways of forming filtrations:

(1) The input is a finite set of points (point cloud) embedded in Euclidean space or other metric space. For any real number  $r > 0$ , consider the union of all spheres of radius  $r$  centered at fixed points in the point cloud. As the radius  $r$  increases, the union of balls provides a filtration.

(2) The input is a real valued function on space  $X$ . For instance,  $X$  can be a Euclidean space. Considering the sublevel set  $\{x \in X \mid f(x) \leq r\}$ , the filtration is obtained as the threshold  $r$  increases.

# Persistent Homology

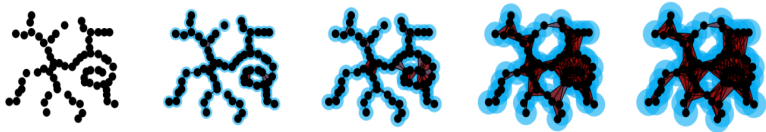


FIGURE 1. A dataset, the surrounding union of balls, and its Čech complexes at different choices of scale.

# Persistent Homology

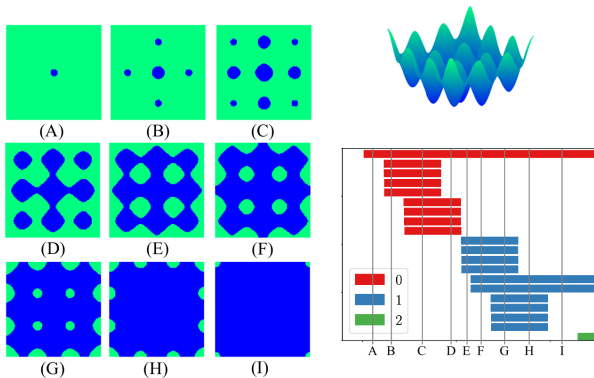


FIGURE 2. (Top right) An energy function for the molecule pentane. The domain is a torus, i.e. a square with periodic boundary conditions, as there are two circular degrees of freedom (dihedral angles) in the molecule. (Left) Nine different sublevel sets of energy. (Bottom right) The sublevel set persistent homology of this energy function on the torus, with 0-dimensional homology in red, 1-dimensional homology in blue, 2-dimensional homology in green. Image from [Mirth et al. \[2020\]](#).

# Persistent Homology

Persistent homology can be represented in two equivalent forms: either as a **persistence barcode** or as a **persistence diagram**;

**Persistence barcode** is a collection of bars whose left end points and right end points are birth and death respectively.

Each interval in the persistence barcode is represented in the persistence diagram by a point in the plane, with its **birth coordinate** on the horizontal axis and with its **death coordinate** on the vertical axis. Since the death of each feature is after its birth, persistence diagram points all lie above the diagonal line  $y = x$ .

# Persistent Homology

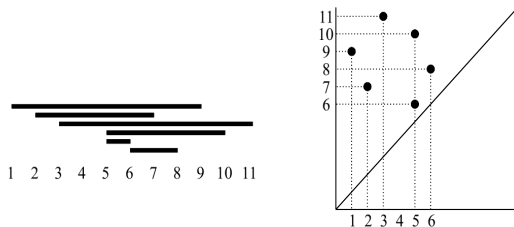


FIGURE 3. (Left) A persistent homology barcode, with the birth and death scale of each bar indicated on the horizontal axis. (Right) Its corresponding persistence diagram, i.e. a collection of points in the first quadrant above the diagonal, with birth coordinates on the horizontal axis and death coordinates on the vertical axis.

# Outline

- 1 Motivation
- 2 Introduction
- 3 Application to Plant Morphology Analysis
- 4 References

# Steps of the PH method

## Steps of the PH method

1. Extracting point cloud datasets from images
2. Converting point clouds to functions
3. Forming filtrations
4. Computing persistent homology
5. Analyzing persisting results(e.g., persistent barcode/diagram, Betti curve and Euler characteristic)

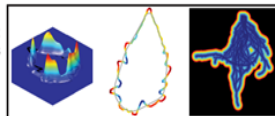
# Steps of the PH method

**B**

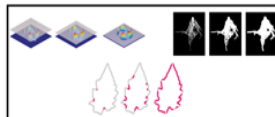
1. Extracting point clouds from images



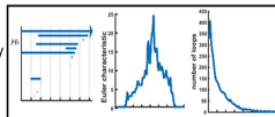
2. Converting point clouds to functions



3. Forming filtrations



4. Computing persistent homology





# Process of analysis

1. Combine PH with statistical framework to diverse morphologies in the same tomato plants of a near-isogenic *S. pennellii* introgression line population (compare the introgression line ILs to the cv M82 parent).
2. Apply statistical methods such as multidimensional scaling (MDS), principal component analysis (PCA), and canonical variant analysis (CVA) to turn the barcode information into a single value that describes leaf shape, leaflet serrations, and root architecture, to demonstrate the versatility of this approach.

# Process of analysis

get leaves/roots datasets → obtain persistent barcodes →  
compute bottleneck distance → MDS → CVA /PCA → get CV1  
→ make comparisons between near-isogenic introgression lines and  
the parent cv M82 → comprehensive analysis

# Results

1. Compared with conventional univariate or multivariate approaches, PH-based phenotyping can significantly increase the ability to **capture phenotypic variation caused by genetic introgressions**, as evidenced by a greater normalized effect size and **the detection of unique QTLs**.
2. Multivariate analyses (whether with conventional traits or PH) always **outperform** univariate traits. Persistent homology reveals novel QTLs not identified using conventional univariate or multivariate traits.
3. Persistent homology detects **concerted changes in shoot and root architecture**.

# Outline

- 1 Motivation
- 2 Introduction
- 3 Application to Plant Morphology Analysis
- 4 References**

# References

Mao Li et al. The Persistent Homology Mathematical Framework Provides Enhanced Genotype-to-Phenotype Associations for Plant Morphology, *Plant Physiology* Aug 2018, 177 (4) 1382-1395;

Mao Li et al. Characterizing 3D inflorescence architecture in grapevine using X-ray imaging and advanced morphometrics: implications for understanding cluster density, *Journal of Experimental Botany*, Volume 70, Issue 21, 1 November 2019, Pages 6261–6276;

Adams H, Moy M. *Topology Applied to Machine Learning: From Global to Local*[J].2021.