Topological Data Analysis

2022–2023

Lecture 6

Persistence Diagrams

21 November 2022

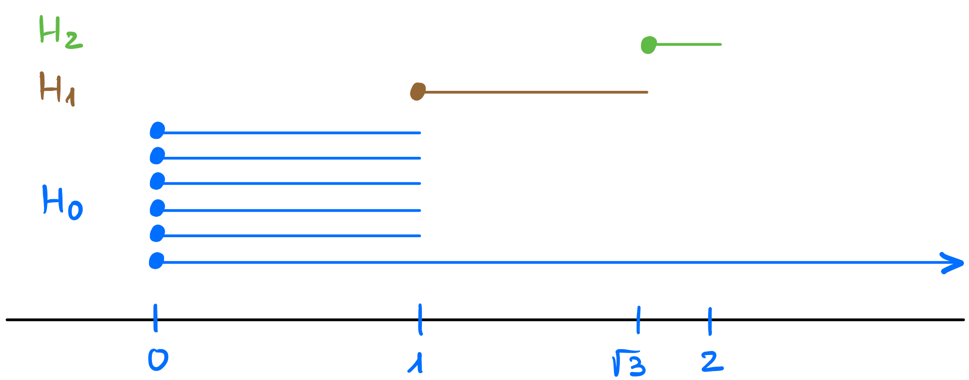
Let (V, π) be a persistence module of finite type over a field $\mathbb{F}$, and let

$$\bigoplus_{i=1}^{n} \mathbb{F}[b_i, d_i) \oplus \bigoplus_{j=1}^{m} \mathbb{F}[c_j, \infty)$$

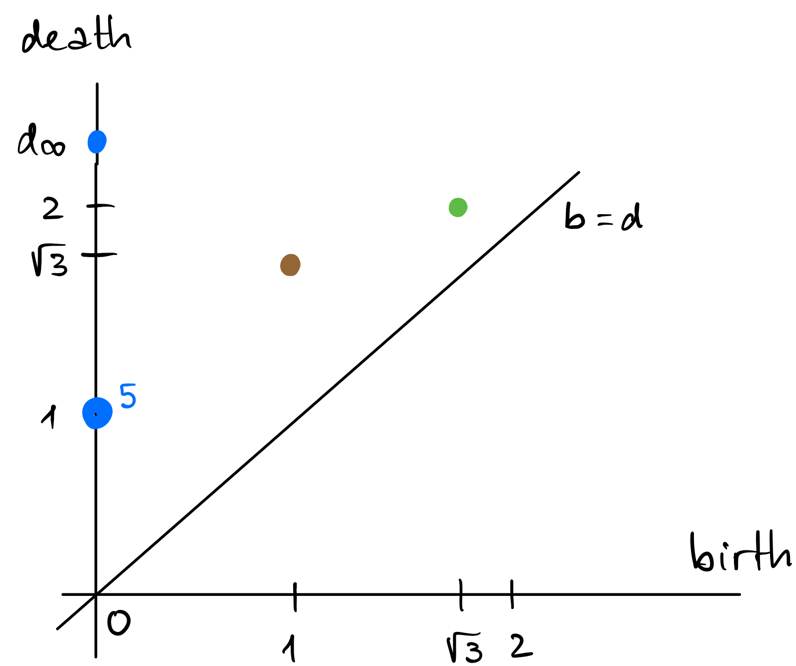
be its normal form.

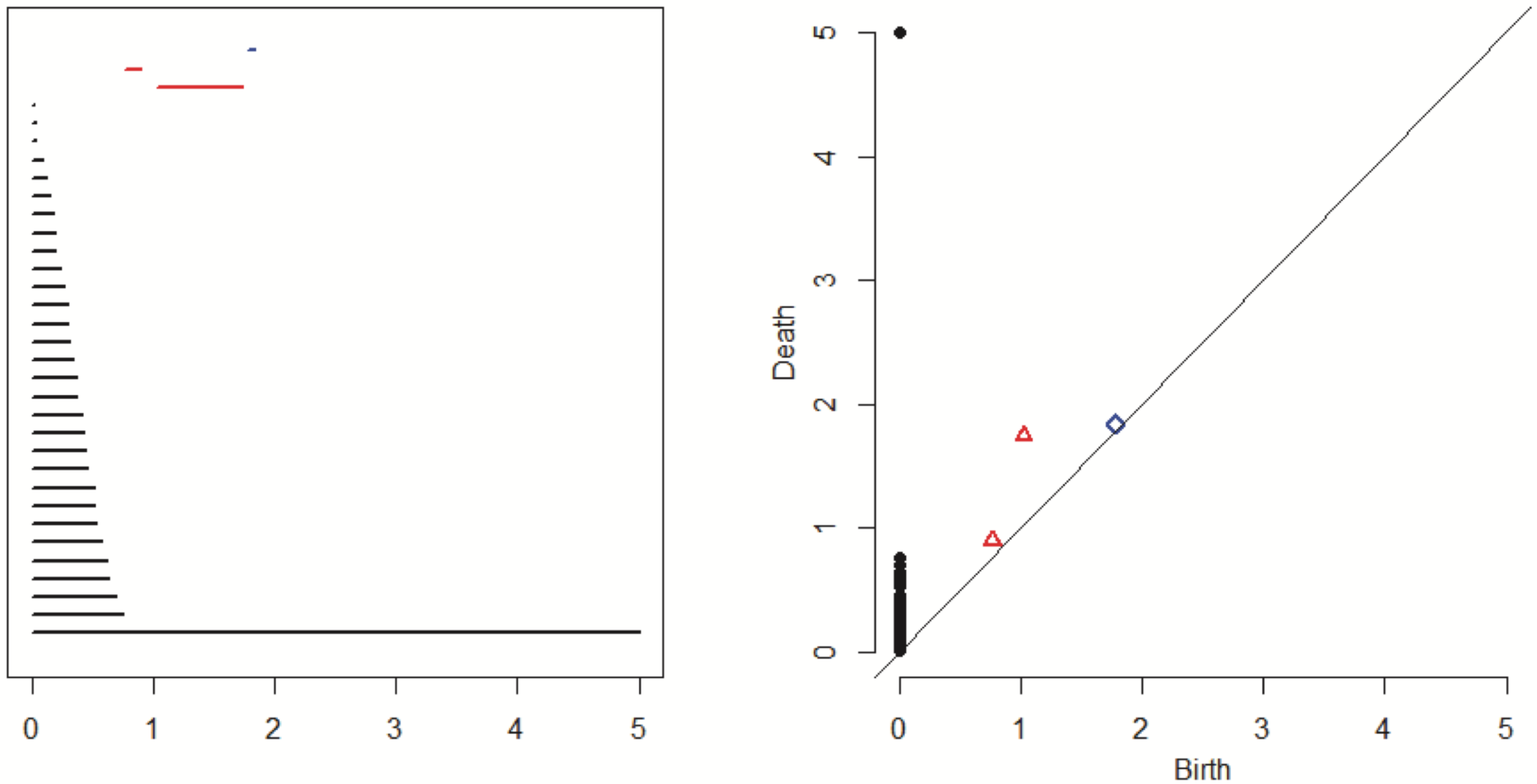
The persistence diagram D of (V, π) has a point (b_i, d_i) for each $i \in \{1, \dots, n\}$ and a point (c_j, d_∞) for each $j \in \{1, \dots, m\}$, where d_∞ is an arbitrary but fixed real number bigger than all values in the spectrum of V.

Multiplicities are depicted with labels on the points of D.

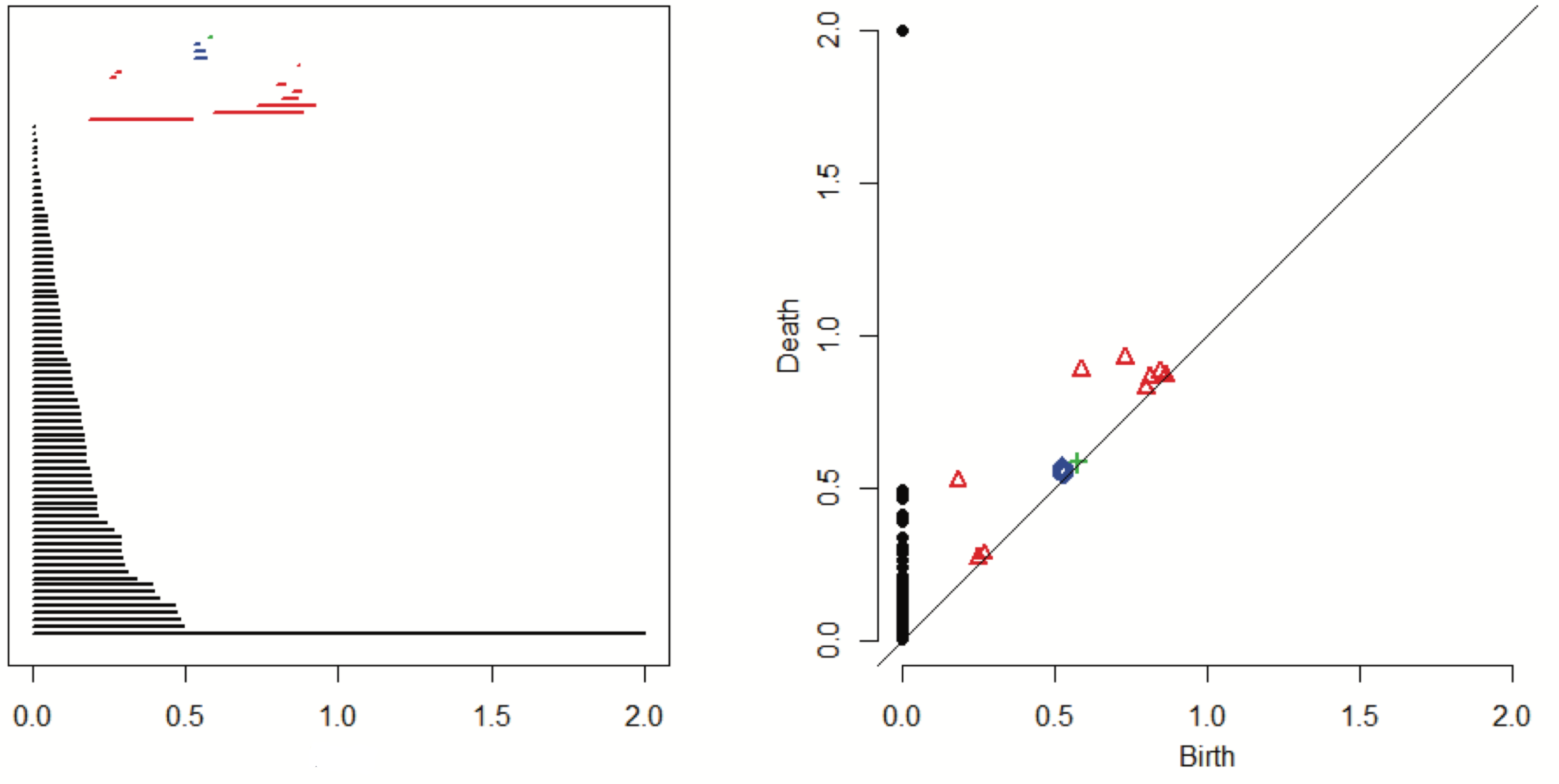


$$\mathbb{F}[0, \infty) \oplus \mathbb{F}[0, 1)^5 \oplus \mathbb{F}[1, \sqrt{3}) \oplus \mathbb{F}[\sqrt{3}, 2)$$





Persistence homology barcode and persistence diagram

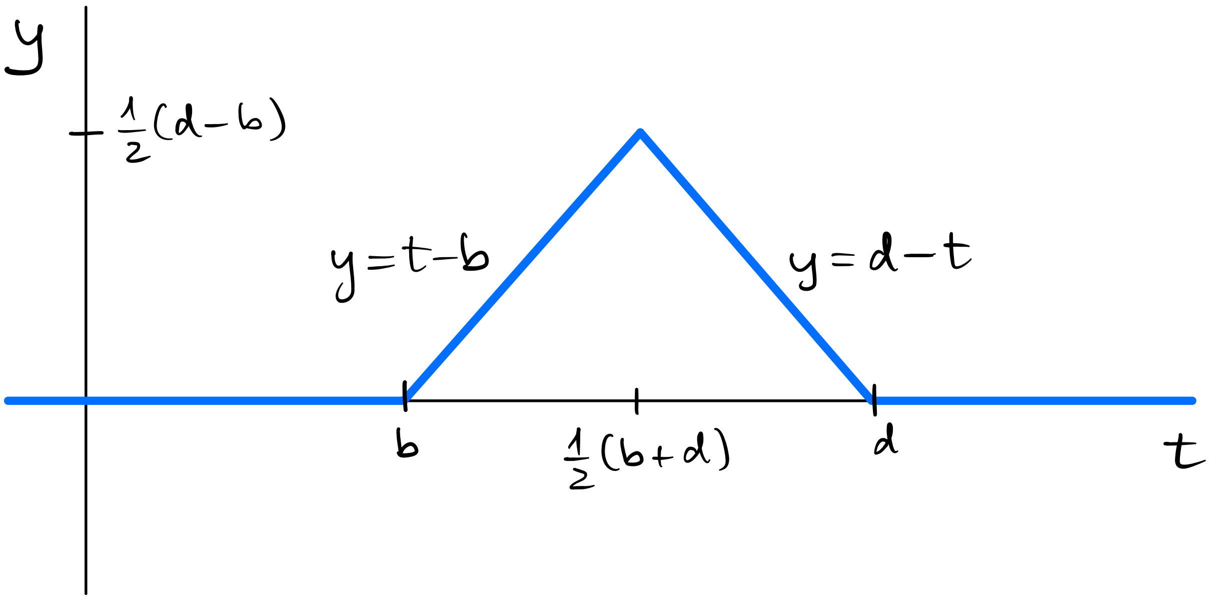


Persistence homology barcode and persistence diagram

Persistence landscapes

For real numbers $b<d$, consider the tent function

$$\Lambda_{(b,d)}(t) = \max\{0, \min\{t-b, d-t\}\}.$$



The landscape of a persistence module (V,π) of finite type is a sequence of piecewise linear functions $\lambda_k : \mathbb{R} \to \mathbb{R}$, $k=0,1,2,\ldots$ defined as follows:

$$\lambda_k(t) = \operatorname*{kmax}_{i\in I}\{\Lambda_{(b_i,d_i)}(t)\}$$

if $\{(b_i,d_i)\}_{i\in I}$ is the multiset of points in the persistence diagram of (V,π) and kmax returns the k^{th} largest value, or zero if there is no such.

A <u>multiset</u> is a set whose elements may have multiplicities. In this case, kmax accounts for multiplicities, if any.

<u>Example:</u>

$1\max\{1,1,1,2,2,5\} = 5$

$2\max\{1,1,1,2,2,5\} = 2$

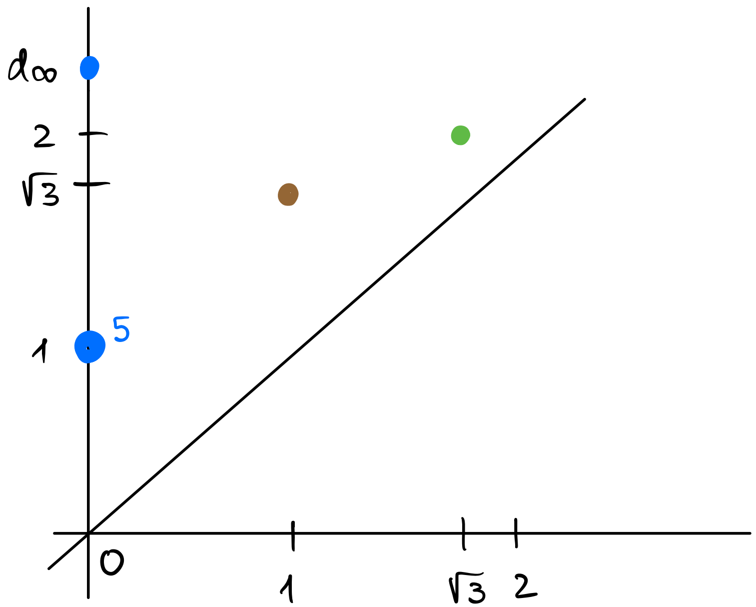
$3\max\{1,1,1,2,2,5\} = 2$

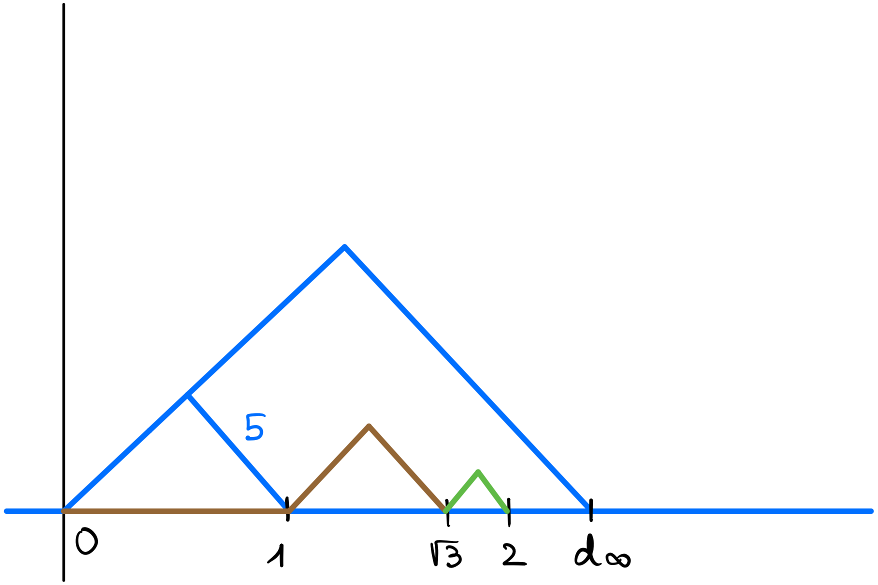
$4\max\{1,1,1,2,2,5\} = 1$

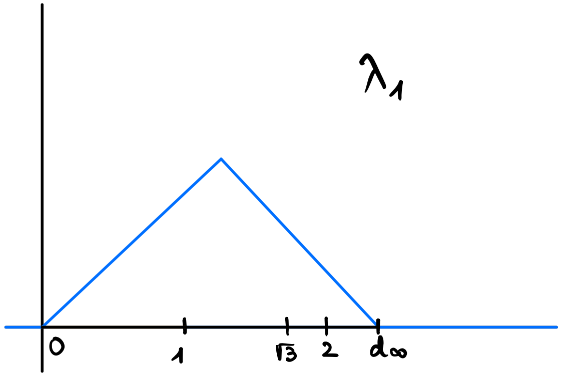
$5\max\{1,1,1,2,2,5\} = 1$

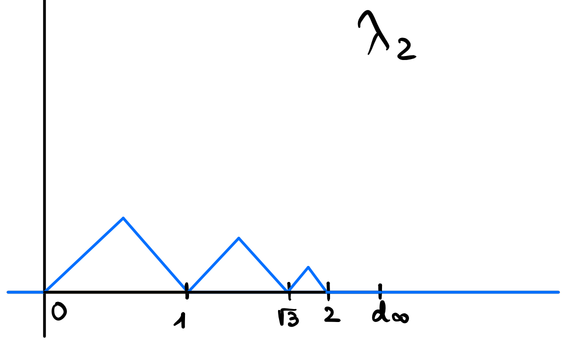
$6\max\{1,1,1,2,2,5\} = 1$

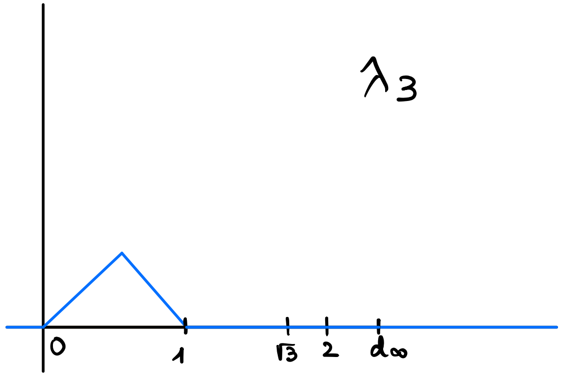
$7\max\{1,1,1,2,2,5\} = 0$

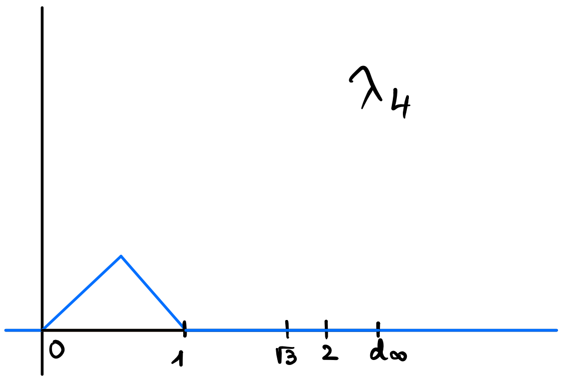


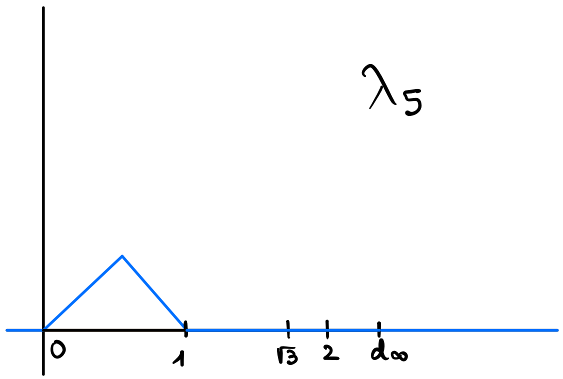


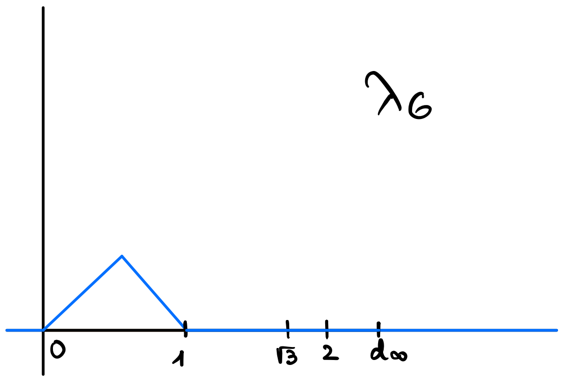

λ_1
0
1
$\sqrt{3}$
2
d_∞

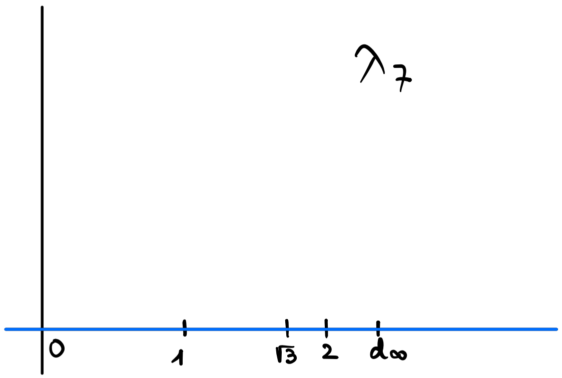

λ_2
0
1
$\sqrt{3}$
2
d_∞

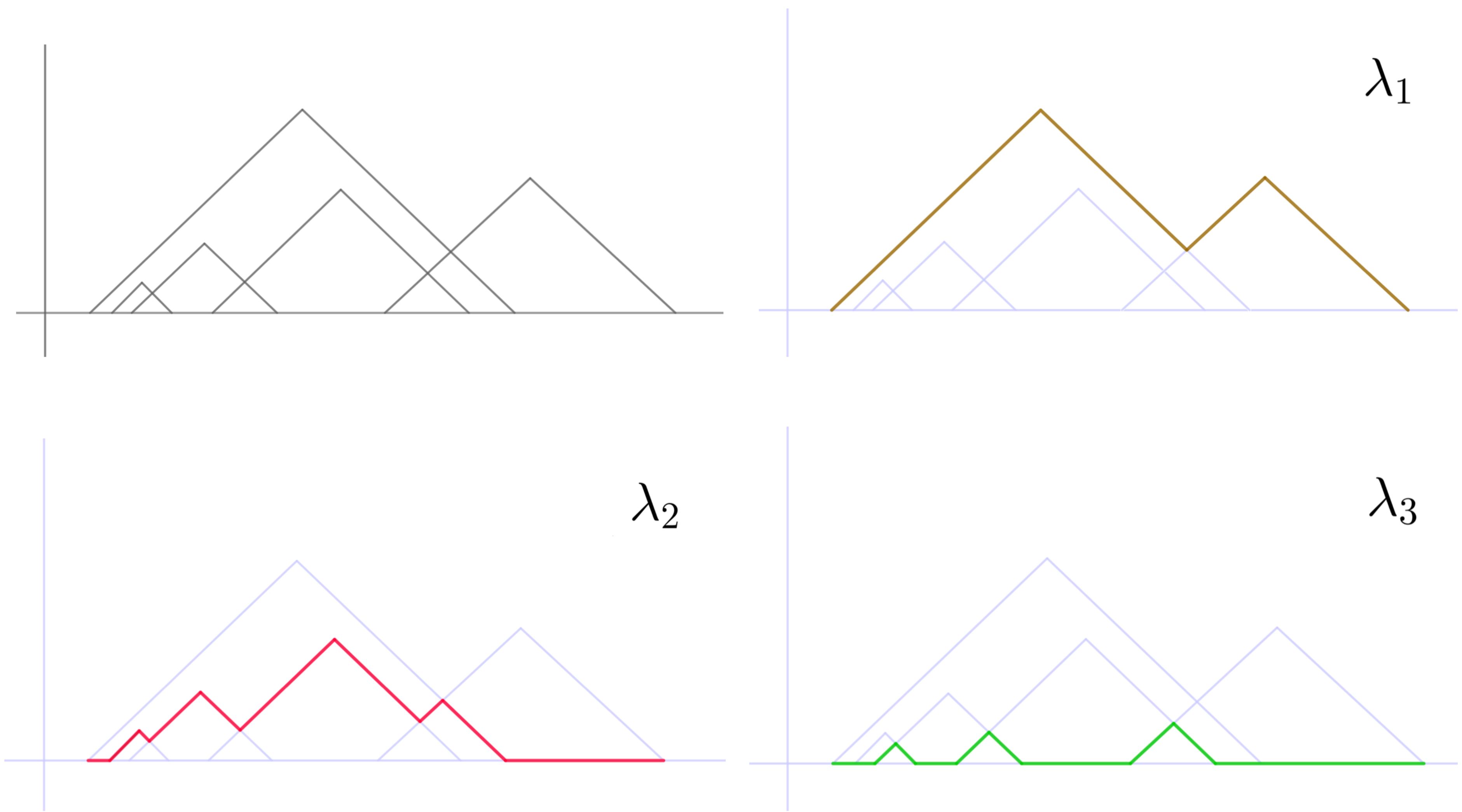

λ_3
0
1
$\sqrt{3}$
2
d_∞


λ_4
0
1
$\sqrt{3}$
2
d_∞

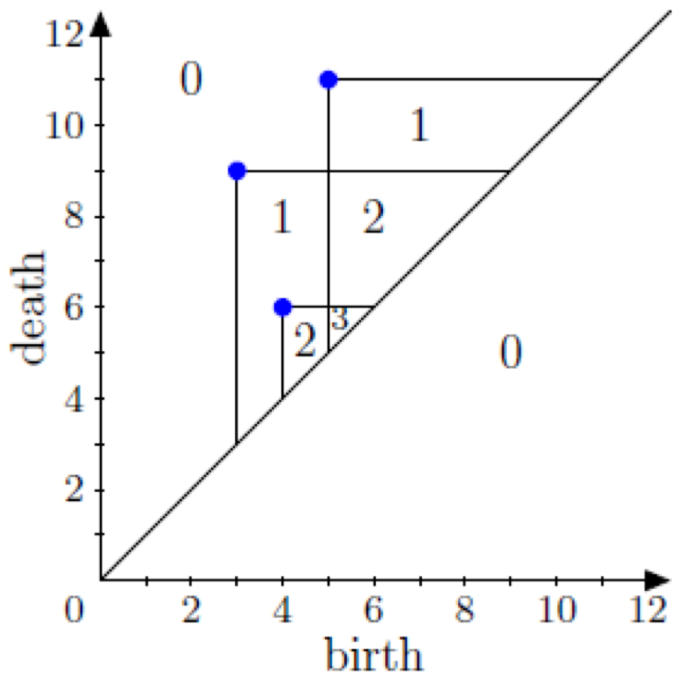

λ_5
0
1
$\sqrt{3}$
2
d_∞

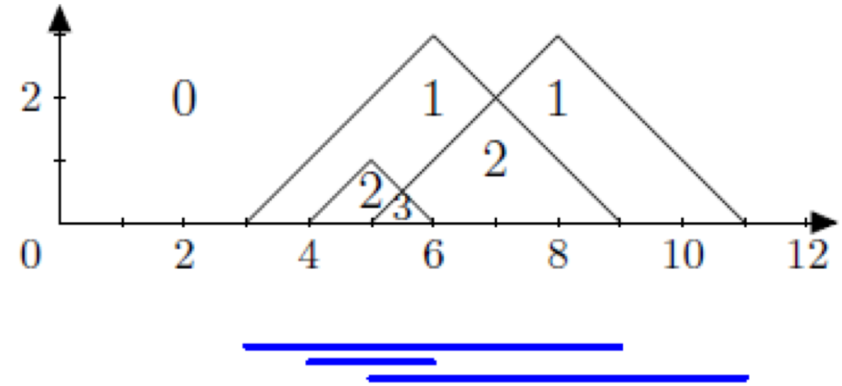

λ_6
0
1
$\sqrt{3}$
2
d_∞

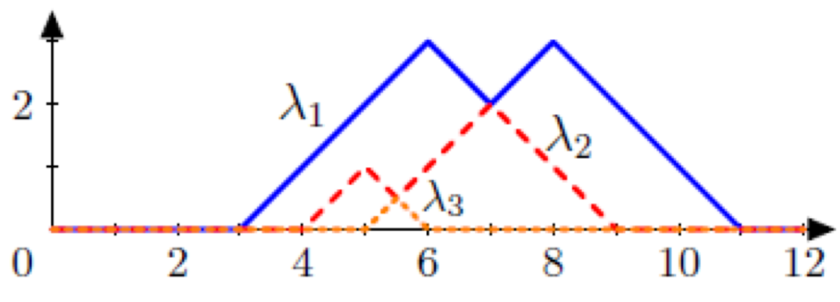

λ_7
0
1
$\sqrt{3}$
2
d_∞

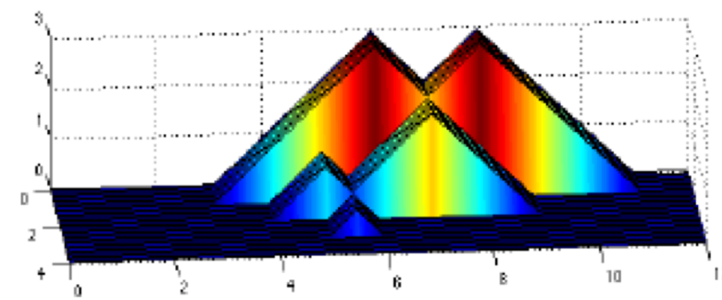


λ_1
λ_2
λ_3



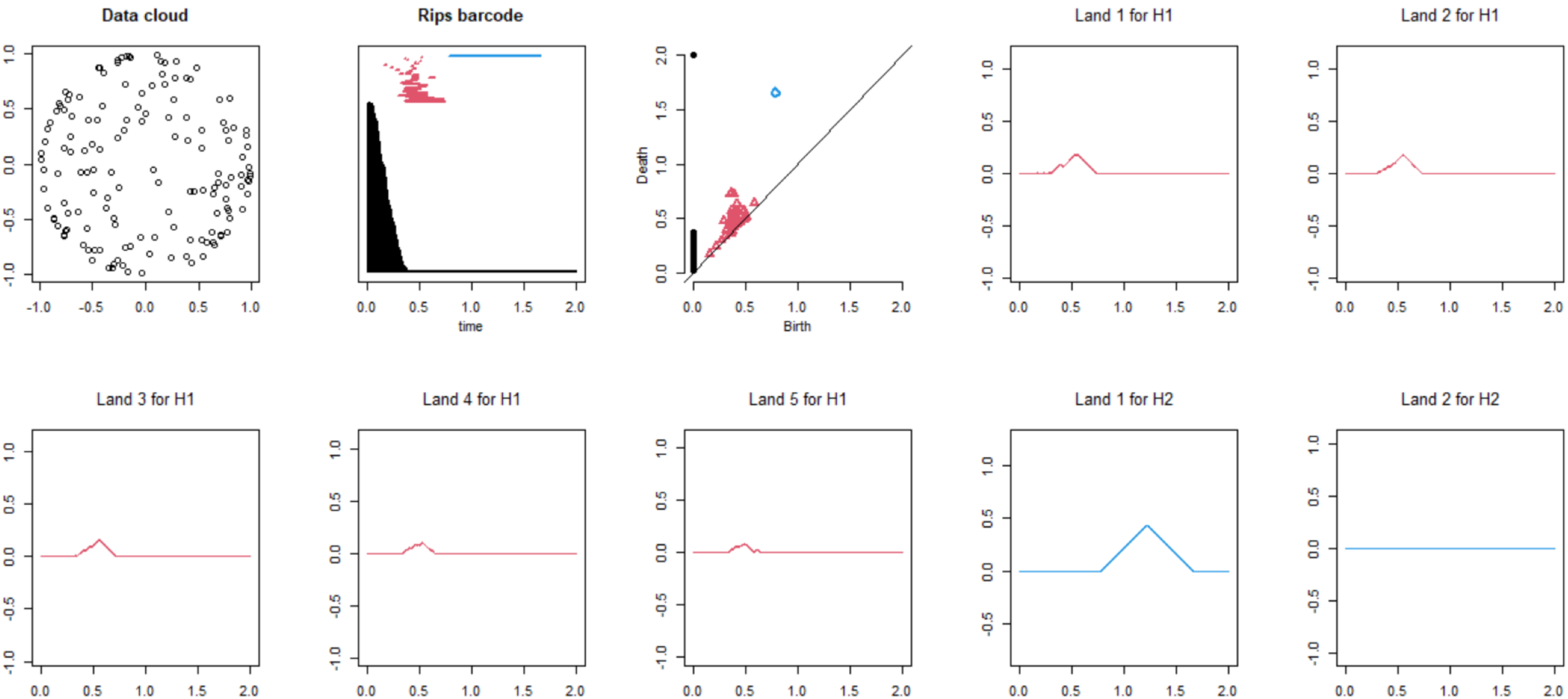






- **P. Bubenik,** Statistical topological data analysis using persistence landscapes, *J. Mach. Learn. Res.* **16** (2015), 77–102, arXiv:1207.6437

https://www.jmlr.org/papers/volume16/bubenik15a/bubenik15a.pdf

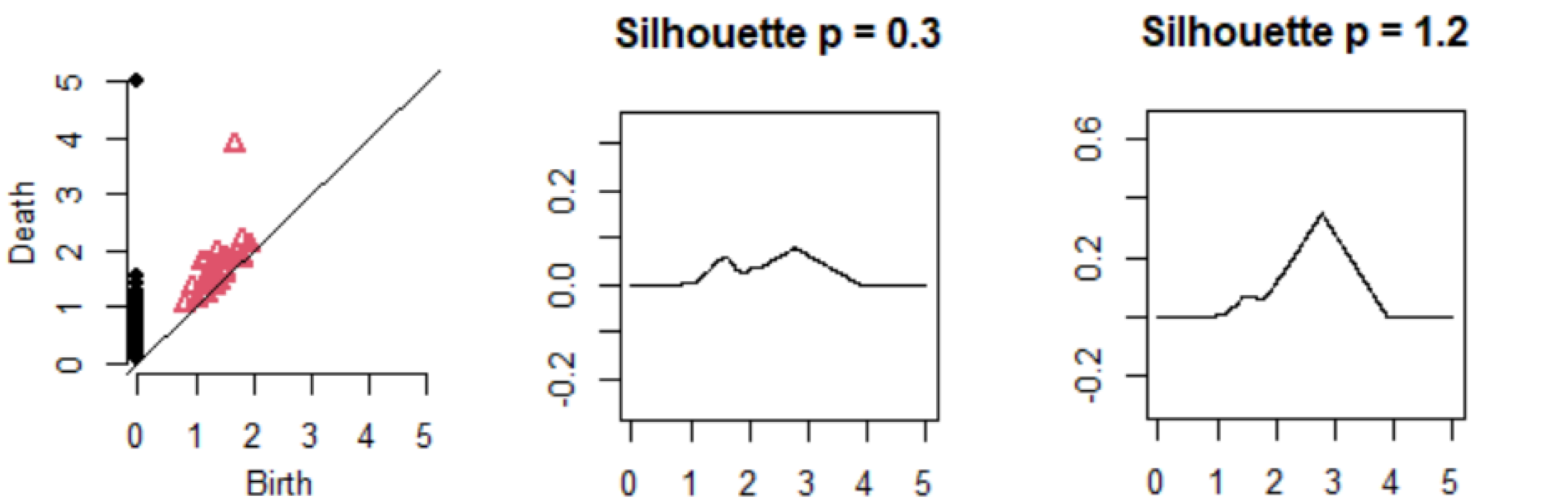

Data cloud
Rips barcode
time
Death
Birth
Land 1 for H1
Land 2 for H1
Land 3 for H1
Land 4 for H1
Land 5 for H1
Land 1 for H2
Land 2 for H2

A silhouette of a persistence diagram $D = \{(b_i, d_i)\}_{i \in I}$ is a weighted average of tent functions from D:

$$\phi(t) = \frac{\sum_i w_i \, \Lambda_{(b_i, d_i)}(t)}{\sum_i w_i}$$

where $w_i \geq 0$ for all i.

A default choice is $w_i = (d_i - b_i)^p$ where p is a parameter. Choosing p small enhances low-persistence features while choosing p large enhances highly persistent features.



- **F. Chazal, B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman,** Stochastic convergence of persistence landscapes and silhouettes, SOCG'14: Proceedings of the Thirtieth Annual Symposium on Computational Geometry, June 2014, pp. 474–483, arXiv:1312.0308

  https://doi.org/10.1145/2582112.2582128