5. Persistence Diagrams

Vin de Silva
Pomona College

CBMS Lecture Series, Macalester College, June 2017

Persistence modules

Persistence modules

A diagram of vector spaces and linear maps

$$V_0 \longrightarrow V_1 \longrightarrow \dots \longrightarrow V_n$$

is called a **persistence module**.

Persistence modules

Persistence modules

A diagram of vector spaces and linear maps

$$V_0 \longrightarrow V_1 \longrightarrow \ldots \longrightarrow V_n$$

is called a **persistence module**.

Example

Let

$$S_0 \xrightarrow{\subseteq} S_1 \xrightarrow{\subseteq} \ldots \xrightarrow{\subseteq} S_n$$

be a nested sequence of simplicial complexes. Then

$$\mathrm{H}_k(S_0) \longrightarrow \mathrm{H}_k(S_1) \longrightarrow \ldots \longrightarrow \mathrm{H}_k(S_n)$$

is a persistence module. The arrows denote the maps in homology induced by the inclusions $S_{k-1} \subseteq S_k$.

Persistence modules

A diagram of vector spaces and linear maps

$$V_0 \longrightarrow V_1 \longrightarrow \dots \longrightarrow V_n$$

is called a **persistence module**.

Example

Let X be a finite metric space and let

$$r_0 < r_1 < \cdots < r_n$$

be the set of values taken by the metric. Then

$$\mathrm{H}_k(\mathrm{VR}(X, r_0)) \longrightarrow \mathrm{H}_k(\mathrm{VR}(X, r_1)) \longrightarrow \dots \longrightarrow \mathrm{H}_k(\mathrm{VR}(X, r_n))$$

is a persistence module. The arrows denote the maps in homology induced by the inclusions $\mathrm{VR}(X, r_{k-1}) \subseteq \mathrm{VR}(X, r_k)$.

Persistence modules

A diagram of vector spaces and linear maps

$$V_0 \longrightarrow V_1 \longrightarrow \dots \longrightarrow V_n$$

is called a **persistence module**.

Example

Let M be a compact smooth manifold and let $f : M \to \mathbb{R}$ be a smooth function with finitely many critical values

$$t_0 < t_1 < \cdots < t_n.$$

Let $M^t = f^{-1}(-\infty, t]$. Then

$$\mathrm{H}_k(M^{t_0}) \longrightarrow \mathrm{H}_k(M^{t_1}) \longrightarrow \dots \longrightarrow \mathrm{H}_k(M^{t_n})$$

is a persistence module. The arrows denote the maps in homology induced by the inclusions $M^{t_{k-1}} \subseteq M^{t_k}$.

Persistence modules

Persistence modules

A **persistence module indexed by** $\mathbb{N}$ is an infinite diagram

$$V_0 \longrightarrow V_1 \longrightarrow \dots \longrightarrow V_n \longrightarrow \dots$$

of vector spaces and linear maps.

Zomorodian, Carlsson (2002)

This is the same thing as a graded module over the polynomial ring $\mathbb{F}[t]$.

Field elements act on each V_k by scalar multiplication. The ring element t acts by applying the map $V_k \to V_{k+1}$.

The rank invariant

n=0

Diagrams of the form

$$V_0$$

are classified up to isomorphism by a single numerical invariant:

$$\dim(V_0)$$

The rank invariant

n=0

Diagrams of the form

$$V_0$$

are classified up to isomorphism by a single numerical invariant:

$$r(0,0) = \mathrm{rank}(V_0 \to V_0) = \dim(V_0)$$

The rank invariant

n=1

Diagrams of the form

$$V_0 \longrightarrow V_1$$

are classified up to isomorphism by three numerical invariants:

$$\dim(V_0)$$
$$\operatorname{rank}(V_0 \to V_1)$$
$$\dim(V_1)$$

Proof

Represent the map by a matrix. Changes of basis of V_0, V_1 correspond to column and row operations. Every matrix can be reduced to the block form:

$$\left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]$$

These block forms are precisely classified by the three invariants.

The rank invariant

n=1

Diagrams of the form

$$V_0 \longrightarrow V_1$$

are classified up to isomorphism by three numerical invariants:

$$\begin{aligned} r(0,0) &= \dim(V_0) = \operatorname{rank}(V_0 \to V_0) \\ r(0,1) & = \operatorname{rank}(V_0 \to V_1) \\ r(1,1) &= \dim(V_1) = \operatorname{rank}(V_1 \to V_1) \end{aligned}$$

Proof

Represent the map by a matrix. Changes of basis of V_0, V_1 correspond to column and row operations. Every matrix can be reduced to the block form:

$$\left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]$$

These block forms are precisely classified by the three invariants.

The rank invariant

Theorem (Zomorodian, Carlsson, 2002)

Diagrams of the form

$$V_0 \longrightarrow V_1 \longrightarrow \dots \longrightarrow V_n$$

are classified up to isomorphism by the invariants

$$r(i,j) = \mathrm{rank}(V_i \to V_j)$$

for all $0 \leq i \leq j \leq n$.

The rank invariant

Theorem (Zomorodian, Carlsson, 2002)

Diagrams of the form

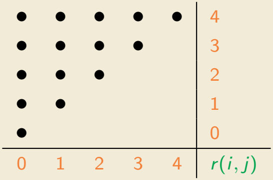
$$V_0 \longrightarrow V_1 \longrightarrow \dots \longrightarrow V_n$$

are classified up to isomorphism by the invariants

$$r(i,j) = \mathrm{rank}(V_i \to V_j)$$

for all $0 \leq i \leq j \leq n$.

Plot these as points-with-multiplicity on a 'rank diagram':



Interval decomposition

Direct sums

Given two persistence modules

$$V_0 \longrightarrow V_1 \longrightarrow \dots \longrightarrow V_n$$

$$W_0 \longrightarrow W_1 \longrightarrow \dots \longrightarrow W_n$$

we can form their direct sum:

$$V_0 \oplus W_0 \longrightarrow V_1 \oplus W_1 \longrightarrow \dots \longrightarrow V_n \oplus W_n$$

Direct sum decomposition

Given a persistence module

$$V_0 \longrightarrow V_1 \longrightarrow \dots \longrightarrow V_n$$

can it be expressed as the direct sum of two non-trivial persistence modules? If it can't, then it is **indecomposable**.

Interval decomposition

Interval modules

The **interval module** $I(i,j)$ is defined to be the persistence module with

$$V_k = \begin{cases} \mathbb{F} & \text{if } i \le k \le j \\ 0 & \text{otherwise} \end{cases}$$

and all maps $\mathbb{F} \to \mathbb{F}$ being the identity (the other maps necessarily being zero).

Here are examples of **interval modules** with $n = 5$:

$$0 \longrightarrow 0 \longrightarrow \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \longrightarrow 0$$

$$0 \longrightarrow 0 \longrightarrow \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \longrightarrow \mathbb{F}$$

$$\mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \longrightarrow \mathbb{F}$$

Interval decomposition

Interval modules

The **interval module** $I(i, j)$ is defined to be the persistence module with

$$V_k = \begin{cases} \mathbb{F} & \text{if } i \leq k \leq j \\ 0 & \text{otherwise} \end{cases}$$

and all maps $\mathbb{F} \to \mathbb{F}$ being the identity (the other maps necessarily being zero).

Here are examples of **interval modules** with $n = 5$:

$$0 \longrightarrow 0 \longrightarrow \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \longrightarrow 0$$

$$0 \longrightarrow 0 \longrightarrow \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \longrightarrow \mathbb{F}$$

$$\mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \xrightarrow{1} \mathbb{F} \longrightarrow \mathbb{F}$$

Fact

Interval modules are indecomposable.

Interval decomposition

Theorem (Gabriel 1970; Zomorodian & Carlsson 2002; etc.)

Every (finite-dimensional) persistence module is isomorphic to a (finite) direct sum of interval modules. The multiplicity

$$m(i, j)$$

of the summand $I(i, j)$ is an invariant of the persistence module.

Interval decomposition

Theorem (Gabriel 1970; Zomorodian & Carlsson 2002; etc.)

Every (finite-dimensional) persistence module is isomorphic to a (finite) direct sum of interval modules. The multiplicity

$$m(i, j)$$

of the summand $I(i, j)$ is an invariant of the persistence module.

Proof

Zomorodian & Carlsson use the classification theorem for finitely-generated graded modules over $\mathbb{F}[t]$.

Interval decomposition

Theorem (Gabriel 1970; Zomorodian & Carlsson 2002; etc.)

Every (finite-dimensional) persistence module is isomorphic to a (finite) direct sum of interval modules. The multiplicity

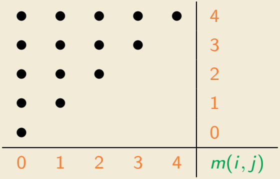
$$m(i, j)$$

of the summand $I(i, j)$ is an invariant of the persistence module.

Proof

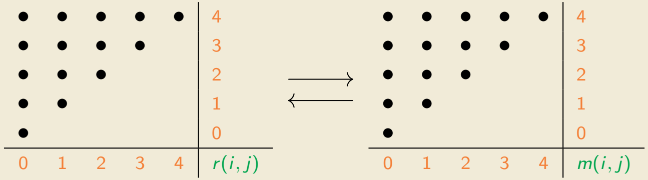
Zomorodian & Carlsson use the classification theorem for finitely-generated graded modules over $\mathbb{F}[t]$.

Plot the multiplicities as points-with-multiplicity on a ‘persistence diagram’:



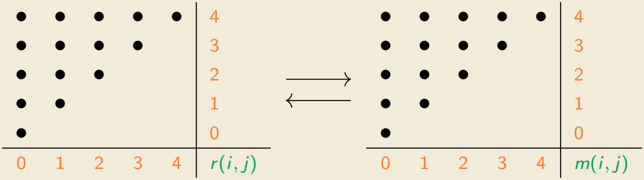
Relationship between the two diagrams

Inverse transformations



Relationship between the two diagrams

Inverse transformations

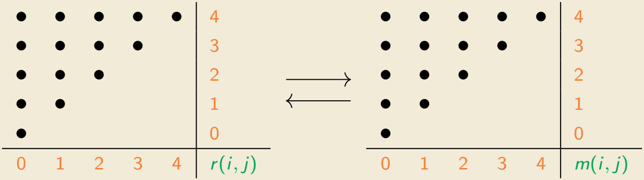


rank diagram ← persistence diagram

Each summand $I(p, q)$ contributes to $r(i, j)$ whenever $p \leq i \leq j \leq q$. Thus:

$$r(i, j) = \sum_{p=0}^{i} \sum_{q=j}^{n} m(p, q)$$

Inverse transformations



rank diagram ← persistence diagram

Each summand $I(p, q)$ contributes to $r(i, j)$ whenever $p \leq i \leq j \leq q$. Thus:

$$r(i,j) = \sum_{p=0}^{i} \sum_{q=j}^{n} m(p,q)$$

rank diagram → persistence diagram

The inverse transformation is

$$m(i,j) = r(i,j) - r(i-1,j) - r(i,j+1) + r(i-1,j+1)$$

by the inclusion exclusion principle.

Relationship between the two diagrams

Inverse transformations

0	0	0	0	0	4
0	2	2	2		3
1	3	3			2
1	3				1
1					0
0	1	2	3	4	$r(i,j)$

$\rightleftarrows$

○	○	○	○	○	4
○	2	○	○		3
1	○	○			2
○	○				1
○					0
0	1	2	3	4	$m(i,j)$

rank diagram $\leftarrow$ persistence diagram

Each summand $I(p,q)$ contributes to $r(i,j)$ whenever $p \leq i \leq j \leq q$. Thus:

$$r(i,j) = \sum_{p=0}^{i} \sum_{q=j}^{n} m(p,q)$$

rank diagram $\rightarrow$ persistence diagram

The inverse transformation is

$$m(i,j) = r(i,j) - r(i-1,j) - r(i,j+1) + r(i-1,j+1)$$

by the inclusion exclusion principle.

Plotting the persistence diagram

Plotting conventions

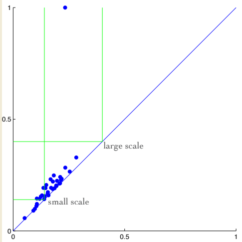
For persistence modules such as the Vietoris–Rips homology of a finite metric space, constructed with respect to a set of critical values

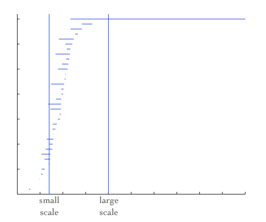
$$r_0 < r_1 < \cdots < r_n$$

we adopt the following convention. Each summand $I(i, j)$ is drawn as

- a half-open interval $[r_i, r_{j+1})$ in the barcode; or
- a point (r_i, r_{j+1}) in the persistence diagram.

We set $r_{n+1} = +\infty$ by convention.





The ELZ Persistence Algorithm

Edelsbrunner, Letscher, Zomorodian (2000)

Given a nested sequence of simplicial complexes

$$S_0 \xrightarrow{\subseteq} S_1 \xrightarrow{\subseteq} \dots \xrightarrow{\subseteq} S_n$$

we wish to compute the persistence diagram of

$$\mathrm{H}_d(S_0) \longrightarrow \mathrm{H}_d(S_1) \longrightarrow \dots \longrightarrow \mathrm{H}_d(S_n).$$

We compute it for all d simultaneously (up to some maximum dimension).

The ELZ Persistence Algorithm

Edelsbrunner, Letscher, Zomorodian (2000)

Given a nested sequence of simplicial complexes

$$S_0 \xrightarrow{\subseteq} S_1 \xrightarrow{\subseteq} \dots \xrightarrow{\subseteq} S_n$$

we wish to compute the persistence diagram of

$$\mathrm{H}_d(S_0) \longrightarrow \mathrm{H}_d(S_1) \longrightarrow \dots \longrightarrow \mathrm{H}_d(S_n).$$

We compute it for all d simultaneously (up to some maximum dimension).

Input

A sequence of 'cells' $\sigma_0, \sigma_1, \dots, \sigma_n$. For each cell we are given:

- the dimension $d = \dim(\sigma_k)$;
- the boundary $\partial[\sigma_k]$, as a cycle built from existing $(d-1)$-cells.

Output

The persistence diagram for this nested sequence of formal cell-complexes.

The ELZ Persistence Algorithm

Data structure

We maintain an upper-triangular square matrix, size equal to the number of cells processed so far. Columns are coloured **green**, **yellow**, **red**.

colour	content	leading coefficient
green	d-cycle α_k	1
yellow	d-cycle β_k	1
red	d-chain γ_k	non-zero

Here d is the dimension of the cell σ_k whose column it is.

We maintain a bijection between the **yellow** and **red** columns.
If **yellow** column k is paired with **red** column ℓ, then $k < \ell$ and

$$\partial\gamma_\ell = \beta_k.$$

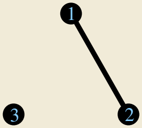
Readout

Each **green** column j corresponds to an interval $[j, +\infty)$.
Each **yellow**–**red** pair (k, ℓ) corresponds to an interval $[k, \ell)$.

The ELZ Persistence Algorithm

Matrix

	1	2	3	12	23	13
1	1	-1				
2		1				
3			1			
12		*		1		
23						
13						



Chains

green : $\alpha_0 = [1],\ \alpha_2 = [3]$

yellow : $\beta_1 = [2] - [1]$

red : $\gamma_3 = [1, 2]$

Persistence intervals

$$[0, +\infty), \quad [1, 3), \quad [2, +\infty)$$

The ELZ Persistence Algorithm

Lemma

The **green** and **yellow** columns form a basis for the space of all cycles.
The **yellow** columns form a basis for the space of all boundaries.

The ELZ Persistence Algorithm

Lemma

The **green** and **yellow** columns form a basis for the space of all cycles.
The **yellow** columns form a basis for the space of all boundaries.

Cycle reduction

Using the table, any cycle ψ can be uniquely written

$$\psi = \partial\rho \quad \text{or} \quad \psi = \partial\rho + \phi$$

ρ is a combination of **red** columns, ϕ is a cycle whose leading simplex is **green**.

The ELZ Persistence Algorithm

Lemma

The **green** and **yellow** columns form a basis for the space of all cycles.
The **yellow** columns form a basis for the space of all boundaries.

Cycle reduction

Using the table, any cycle ψ can be uniquely written

$$\psi = \partial\rho \quad \text{or} \quad \psi = \partial\rho + \phi$$

ρ is a combination of **red** columns, ϕ is a cycle whose leading simplex is **green**.

Iterative method

Initialise $\rho = 0$.
Consider $\psi - \partial\rho$.

- If it is zero then STOP.
- If its leading simplex is **green**, then set $\phi = \psi - \partial\rho$ and STOP.
- If its leading simplex is **yellow**, then add to ρ the appropriate multiple of the associated **red** column to eliminate that leading term.

Repeat.

The ELZ Persistence Algorithm

Update Step

Consider the next simplex σ_k.

The ELZ Persistence Algorithm

Update Step

Consider the next simplex σ_k.

If $\partial[\sigma_k]$ was already a boundary: The cycle reduction lemma gives

$$\partial[\sigma_k] = \partial\rho.$$

We get a new **green** cycle

$$\alpha_k = [\sigma_k] - \rho.$$

The ELZ Persistence Algorithm

Update Step

Consider the next simplex σ_k.

If $\partial[\sigma_k]$ was already a boundary: The cycle reduction lemma gives

$$\partial[\sigma_k] = \partial\rho.$$

We get a new **green** cycle

$$\alpha_k = [\sigma_k] - \rho.$$

If $\partial[\sigma_k]$ was not previously a boundary: The cycle reduction lemma gives

$$\partial[\sigma_k] = \partial\rho + \phi$$

where the leading term of ϕ is $c[\sigma_j]$ for some scalar c.
Change the j-th column to **yellow**, and colour the k-th column **red**. Set

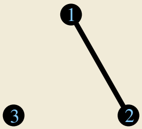
$$\beta_j = c^{-1}\phi, \quad \gamma_k = c^{-1}([\sigma_k] - \rho)$$

(discarding α_j). Pair the columns j and k.

The ELZ Persistence Algorithm

Matrix

	1	2	3	12	23	13
1	1	-1				
2		1				
3			1			
12		*		1		
23						
13						



Chains

green : $\alpha_0 = [1],\ \alpha_2 = [3]$

yellow : $\beta_1 = [2] - [1]$

red : $\gamma_3 = [1, 2]$

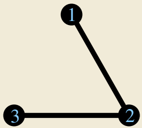
Persistence intervals

$$[0, +\infty), \quad [1, 3), \quad [2, +\infty)$$

The ELZ Persistence Algorithm

Matrix

	1	2	3	12	23	13
1	1	-1				
2		1	-1			
3			1			
12		*		1		
23			*		1	
13						



Chains

green : $\alpha_0 = [1]$

yellow : $\beta_1 = [2] - [1], \ \beta_2 = [3] - [2]$

red : $\gamma_3 = [1,2], \ \gamma_4 = [2,3]$

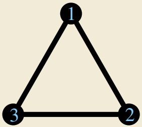
Persistence intervals

$$[0, +\infty), \quad [1,3), \quad [2,4)$$

The ELZ Persistence Algorithm

Matrix

	1	2	3	12	23	13
1	1	-1				
2		1	-1			
3			1			
12		*		1		-1
23			*		1	-1
13						1



Chains

green : $\alpha_0 = [1],\ \alpha_5 = [1,3] - [2,3] - [1,2]$

yellow : $\beta_1 = [2] - [1],\ \beta_2 = [3] - [2]$

red : $\gamma_3 = [1,2],\ \gamma_4 = [2,3]$

Persistence intervals

$$[0, +\infty), \quad [1, 3), \quad [2, 4), \quad [5, +\infty)$$