Computing the rank invariant for persistence bimodules

Jingyi Li

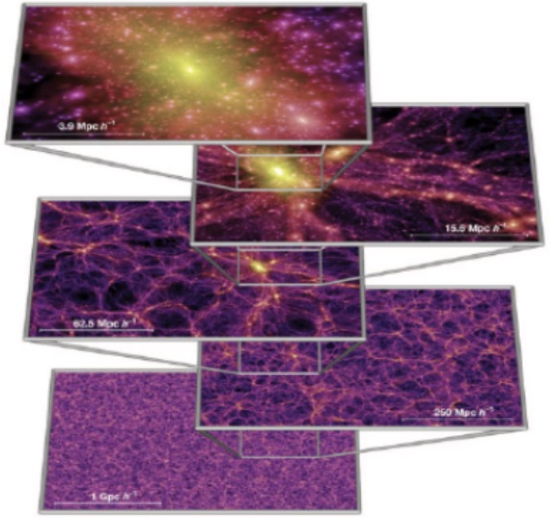
Outline

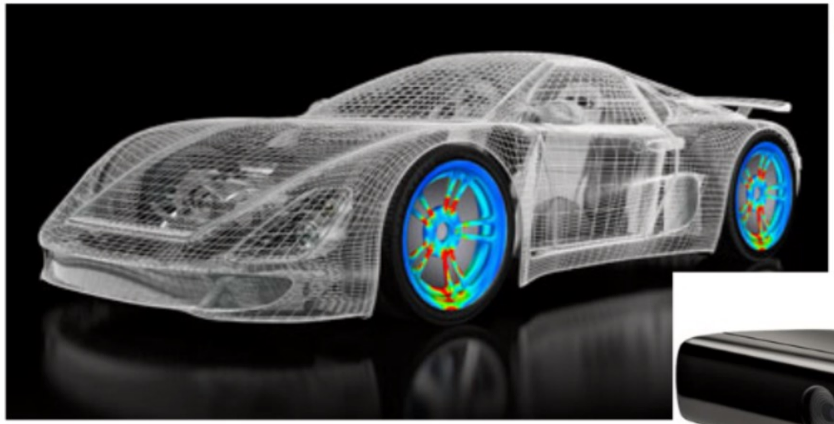
- Introduction
- Background
  - Single-parameter persistence
  - Bipersistence
- Algorithm
  - Description
  - Implementation
  - Complexity
- Experimental results
  - Correctness validation
  - Running time

Context: the data deluge

Data are generated at an unprecedented rate by:

- academia
- industry
- general public

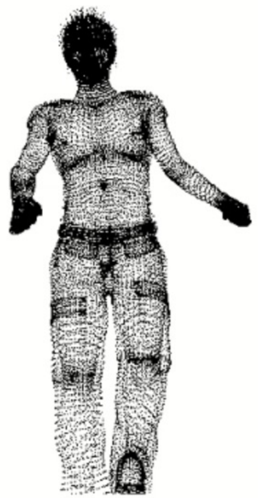


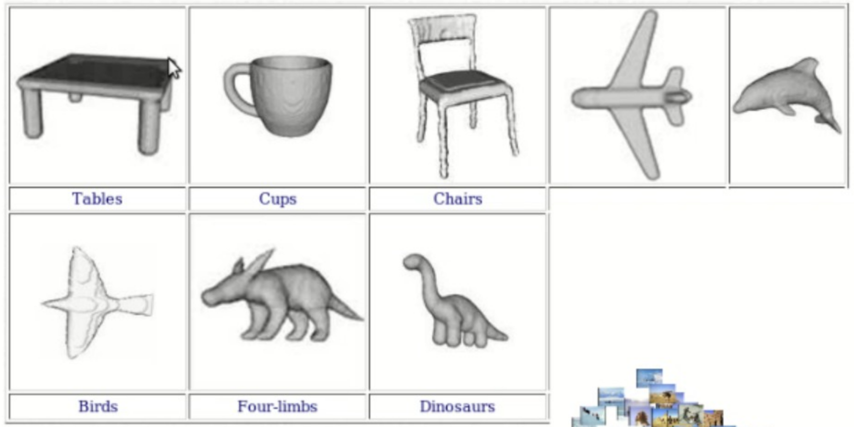


Exploratory analysis of geometric data

Input: set of data points with metric or (dis-)similarity measure

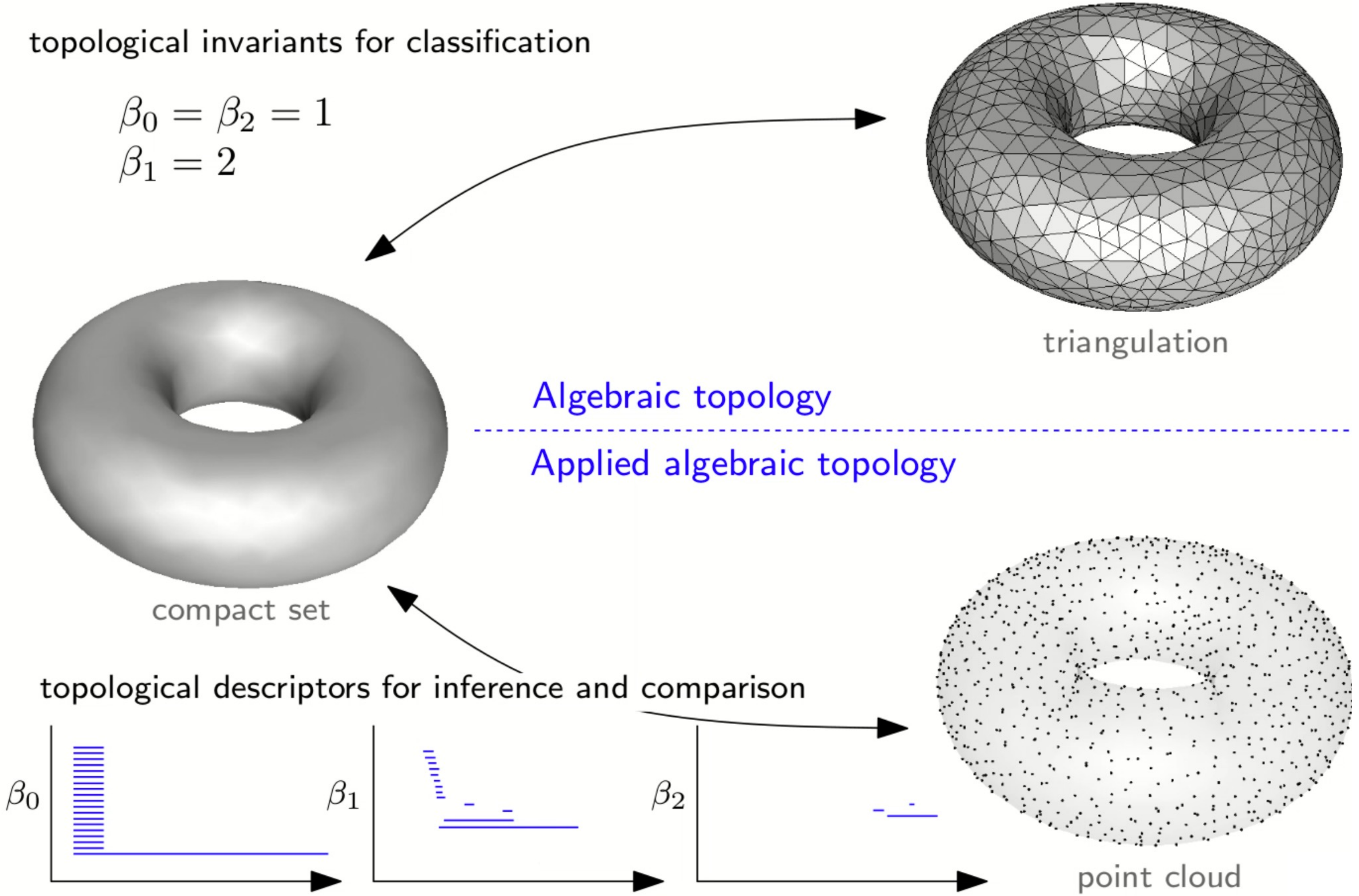
data point $\equiv$ 3d point, image patch, image or 3d shape in collection, Facebook user, etc.





Goal: describe the underlying structure of the data, for interpretation or summary

Topological data analysis (TDA)



Outline

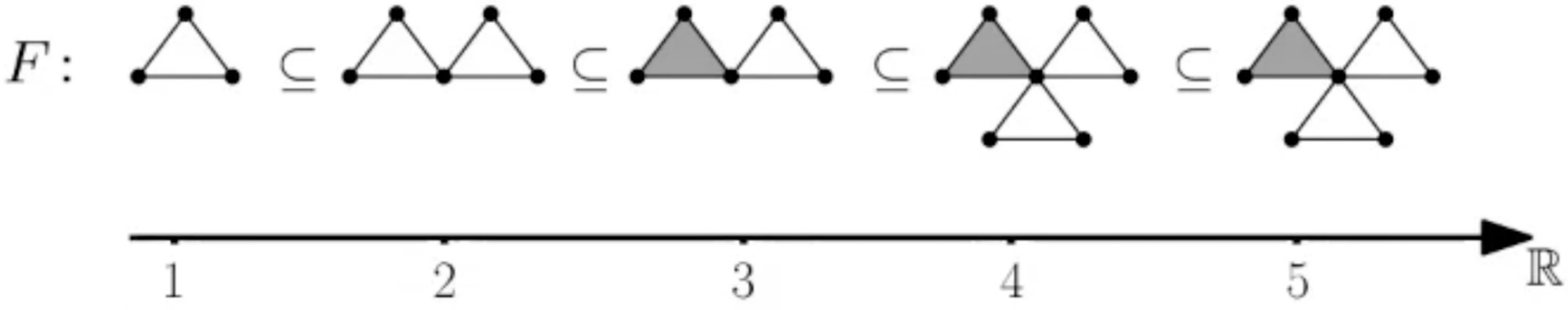
- Introduction
- Background
  - Single-parameter persistence
  - Bipersistence
- Algorithm
  - Description
  - Implementation
  - Complexity
- Experimental results
  - Correctness validation
  - Running time

Filtration

Definition 2.2 (Filtration). *A filtration over U is a family $\mathcal{F} = (F_t)_{t \in U}$ of increasing (for inclusion) topological spaces:*

$$\forall s, t \in U, s \leq t \Rightarrow F_s \subset F_t$$

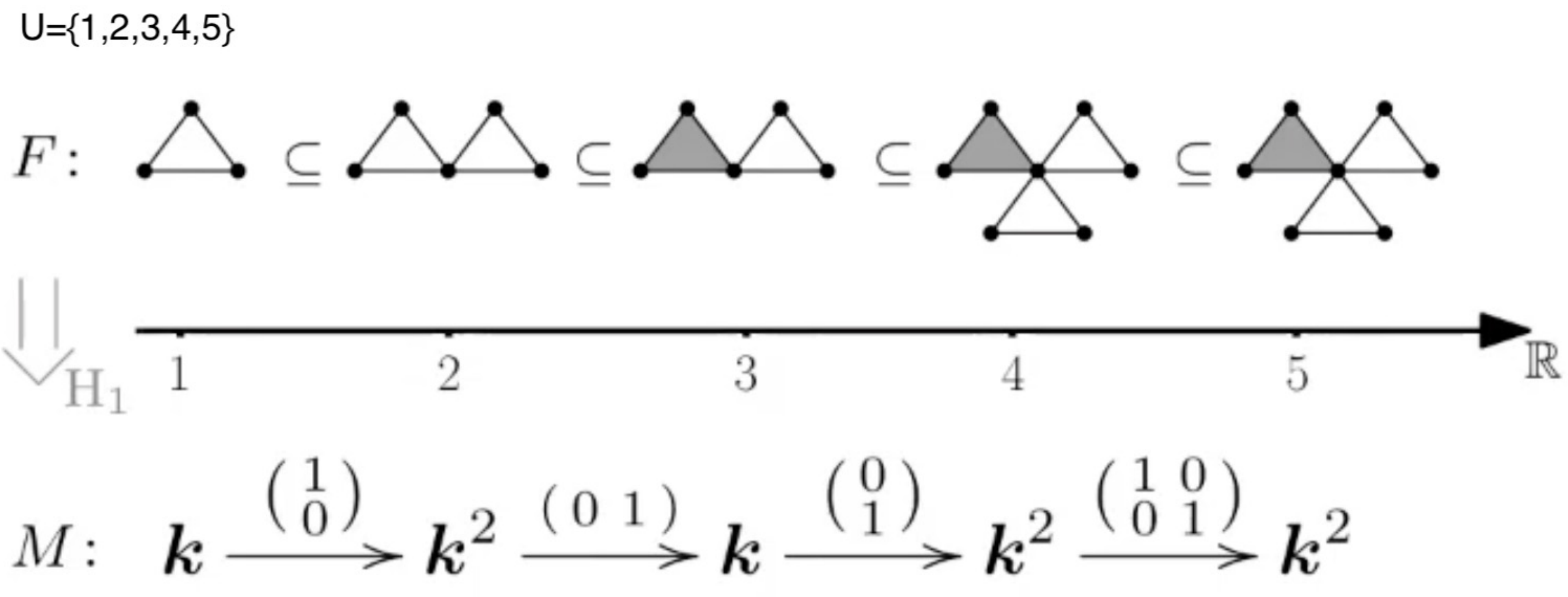
Example: U={1,2,3,4,5}



A persistence module from the filtration

Definition 2.3. *Let $\mathcal{F} = (F_t)_{t \in U}$ be a filtration. We apply the homology functor H_*:*

- $\forall t \in U$, *we define* $V_t := H_*(F_t, \boldsymbol{k})$
- $\forall s \leq t$, *let* v_s^t *be the linear map induced by the canonical inclusion* $F_s \overset{i}{\hookrightarrow} F_t$



Single-parameter persistence

Let $U \subseteq \mathbb{R}$. Let $\boldsymbol{k}$ be a field.

Definition 2.1 (Single-parameter persistence module). *A persistence module $\mathbb{V}$ over U is a collection of $\boxed{\text{vector spaces}}$ $\{V_t\}_{t \in U}$ and $\boxed{\text{linear maps}}$ $v_s^t := V(s \leq t) : V_s \to V_t$ with the property that v_s^s is the identity map and $v_t^u \circ v_s^t = v_s^u$ for all $s \leq t \leq u \in U$.*

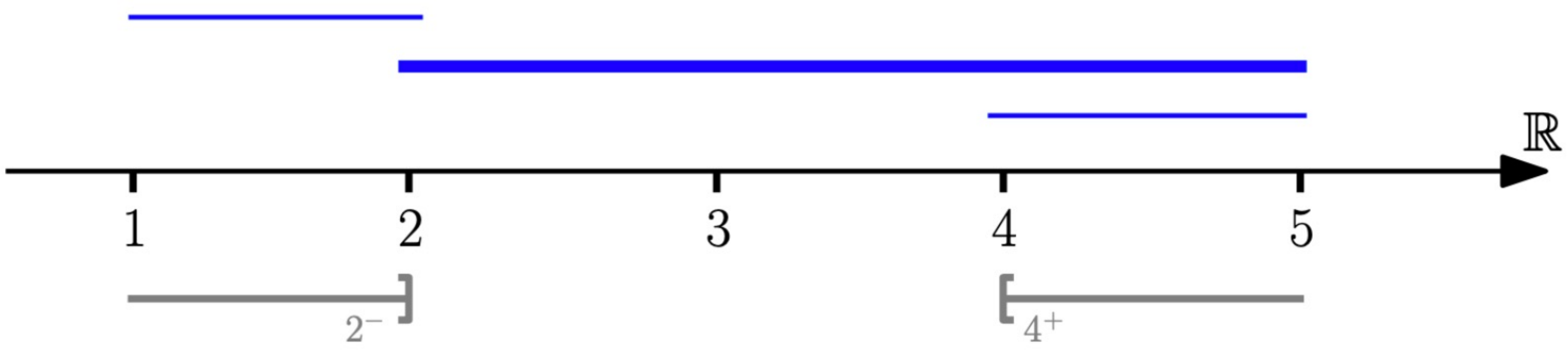
- Vector spaces: $\{V_t\}_{t \in U}$
- Linear maps: $v_s^t = V(s \leq t) \colon V_s \to V_t$

$$v_s^s\colon\ V_s \to V_s = id_{V_s}$$
$$v_t^u \circ v_s^t = v_s^u\colon\ V_s \to V_t \to V_u = V_s \to V_u$$

Decomposition theorem

Theorem 2.1. *A single-parameter persistence module $\mathbb{V}$ can be uniquely decomposed as a direct sum of interval modules, written as: $\mathbb{V} \simeq \bigoplus \mathbb{I}_{[b_j,d_j]}$, when all the vector spaces in $\mathbb{V}$ are finite-dimensional. The multi-set of intervals in the decomposition is called the (persistence) barcode of $\mathbb{V}$, denoted as Dgm $\mathbb{V}$.*

$$\mathbf{k} \xrightarrow{\binom{1}{0}} \mathbf{k}^2 \xrightarrow{(0\ 1)} \mathbf{k} \xrightarrow{\binom{0}{1}} \mathbf{k}^2 \xrightarrow{\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}} \mathbf{k}^2$$



Direct sum of modules

Definition 2.6 (Direct sum of modules). *Given* $\mathbb{V} = (V_t)_t$, $\mathbb{W} = (W_t)_t$ *two persistence modules with corresponding linear maps* (v_s^t), (w_s^t), *we define* $\mathbb{V} \oplus \mathbb{W}$ *by*

- $\mathbb{V} \oplus \mathbb{W} = (V_t \oplus W_t)_t$
- *The corresponding linear maps are denoted by:*

$$(v \oplus w)_s^t : V_s \oplus W_s \to V_t \oplus W_t$$

$$(x, y) \mapsto (v_s^t(x), w_s^t(y))$$

This definition extends naturally to any family of persistence modules $(\mathbb{V}_j)_{j \in J}$, denoted by $\mathbb{V} := \oplus_{j \in J} \mathbb{V}_j$.

Interval Module

Definition 2.4 (Interval). *An interval of U is a convex subset $I \subset U$, i.e. $\forall s \leq t \leq u \in U, (s,u) \in I^2 \Rightarrow t \in I$.*

Definition 2.5 (Interval module). *An interval module over $I \subset U$ is a persistence module $\mathbb{V}$ defined by:*

- *$V_t = \boldsymbol{k}$ if $t \in I$. $V_t = \{0\}$ otherwise.*
- *$\forall s,t \in U, v_s^t = id_{\boldsymbol{k}}$ if $(s,t) \in I^2$. $v_s^t = 0$ otherwise.*

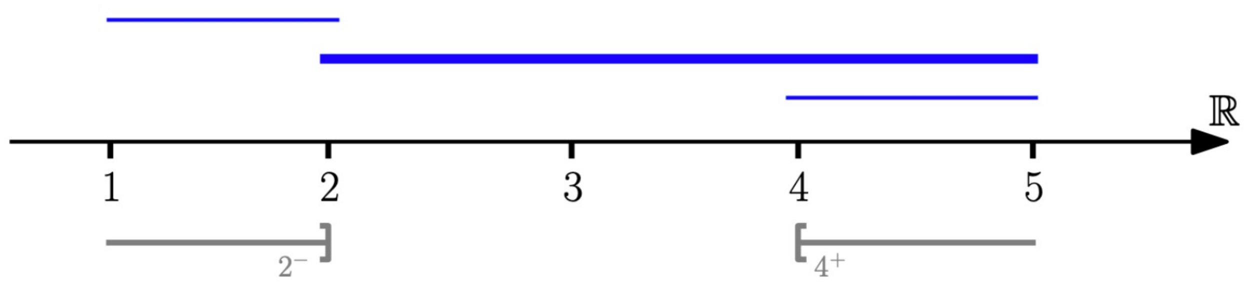
For example, if I=[b,d], the interval module is

$$\mathbb{I}_{[b,d]} := \underbrace{\{0\} \xrightarrow{0} \cdots \xrightarrow{0} \{0\}}_{t<b} \xrightarrow{0} \underbrace{\mathbf{k} \xrightarrow{id} \cdots \xrightarrow{id} \mathbf{k}}_{b \leq t \leq d} \xrightarrow{0} \underbrace{\{0\} \xrightarrow{0} \cdots \xrightarrow{0} \{0\}}_{d<t}$$

By convention, we set $\mathbb{I}_{\emptyset} = 0$.

Multiplicity

$$\mathbf{k} \xrightarrow{\begin{pmatrix}1\\0\end{pmatrix}} \mathbf{k}^2 \xrightarrow{\begin{pmatrix}0 & 1\end{pmatrix}} \mathbf{k} \xrightarrow{\begin{pmatrix}0\\1\end{pmatrix}} \mathbf{k}^2 \xrightarrow{\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}} \mathbf{k}^2$$



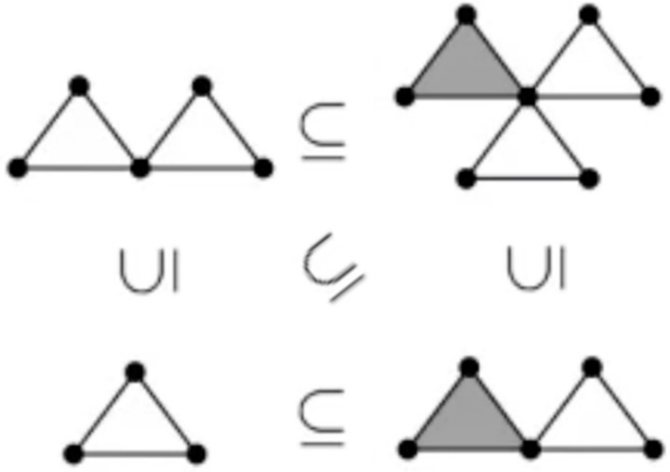
r(2,4)=1
r(4,5)=2

Theorem 2.2 (Multiplicity). *When $U = \mathbb{Z}$, the multiplicity of interval $[s,t] \in Dgm\mathbb{V}$ is given by the following inclusion-exclusion formula:*

$$m(s,t) = r(s,t) - r(s-1,t) - r(s,t+1) + r(s-1,t+1). \tag{2}$$

Definition 2.7 (Rank invariant). *Given a persistence module $\mathbb{V}$ over U, the rank invariant of $\mathbb{V}$ is the collection of ranks $r(s,t) = rank(V(s \leq t))$ for all $s,t \in U$ such that $s \leq t$.*

Bifiltration



Definition 2.9 (Bifiltration). *A bifiltration over U is a family $\mathcal{F} = (F_t)_{t \in U}$ of increasing (for inclusion) topological spaces:*

$$\forall s, t \in U, s \leq t \Rightarrow F_s \subset F_t$$

Definition 2.10. *Let $\mathcal{F} = (F_t)_{t \in U}$ be a bifiltration. We apply the homology functor H_*:*

- *$\forall t \in U$, we define $V_t := H_*(F_t, \boldsymbol{k})$*
- *$\forall s \leq t$, let v_s^t be the linear application induced by the canonical inclusion $F_s \overset{i}{\hookrightarrow} F_t$*

Bipersistence

Let U be a subset of $\mathbb{R}^2$ equipped with the product order: $s \leq t$ iff $s.x \leq t.x$ and $s.y \leq t.y$.

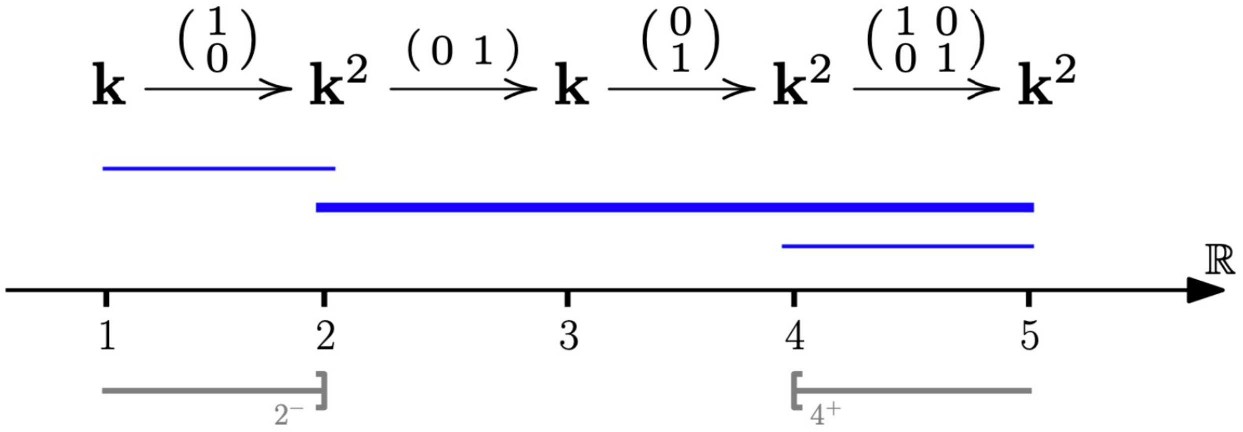
Definition 2.8 (Bimodule). *A bimodule $\mathbb{V}$ over U is a collection of vector spaces $\{V_t\}_{t \in U}$ and linear maps $v_s^t := V(s \leq t) : V_s \to V_t$ with the property that v_s^s is the identity map and $v_t^u \circ v_s^t = v_s^u$ for all $s \leq t \leq u \in U$, where $s \leq t$ if and only if $s.x \leq t.x$ and $s.y \leq t.y$.*

Definition 2.1 (Single-parameter persistence module). *A persistence module $\mathbb{V}$ over U is a collection of vector spaces $\{V_t\}_{t \in U}$ and linear maps $v_s^t := V(s \leq t) : V_s \to V_t$ with the property that v_s^s is the identity map and $v_t^u \circ v_s^t = v_s^u$ for all $s \leq t \leq u \in U$.*

Bipersistence

$$
\begin{tikzcd}
\mathbf{k} \arrow[r, "id"] & \mathbf{k} \arrow[r] & 0 \\
\mathbf{k} \arrow[u, "id"] \arrow[r, "\begin{bmatrix}1\\0\end{bmatrix}"] & \mathbf{k}^2 \arrow[u, "{[1\ 0]}"'] \arrow[r, "{[1\ 1]}"'] & \mathbf{k} \arrow[u] \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u, "\begin{bmatrix}0\\1\end{bmatrix}"] \arrow[r, "id"] & \mathbf{k} \arrow[u, "id"']
\end{tikzcd}
\qquad \cong \qquad \bigoplus \quad ?
$$

Multiplicity



r(2,4)=1
r(4,5)=2

Theorem 2.2 (Multiplicity). *When $U = \mathbb{Z}$, the multiplicity of interval $[s,t] \in Dgm\mathbb{V}$ is given by the following inclusion-exclusion formula:*

$$m(s,t) = r(s,t) - r(s-1,t) - r(s,t+1) + r(s-1,t+1). \tag{2}$$

Definition 2.7 (Rank invariant). *Given a persistence module $\mathbb{V}$ over U, the rank invariant of $\mathbb{V}$ is the collection of ranks $r(s,t) = rank(V(s \leq t))$ for all $s,t \in U$ such that $s \leq t$.*

$$Rk\mathbb{V} = Rk\Big(\bigoplus_{I \in Dgm\mathbb{V}} \mathbf{k}_I\Big) = \sum_{I \in Dgm\mathbb{V}} Rk\mathbf{k}_I.$$

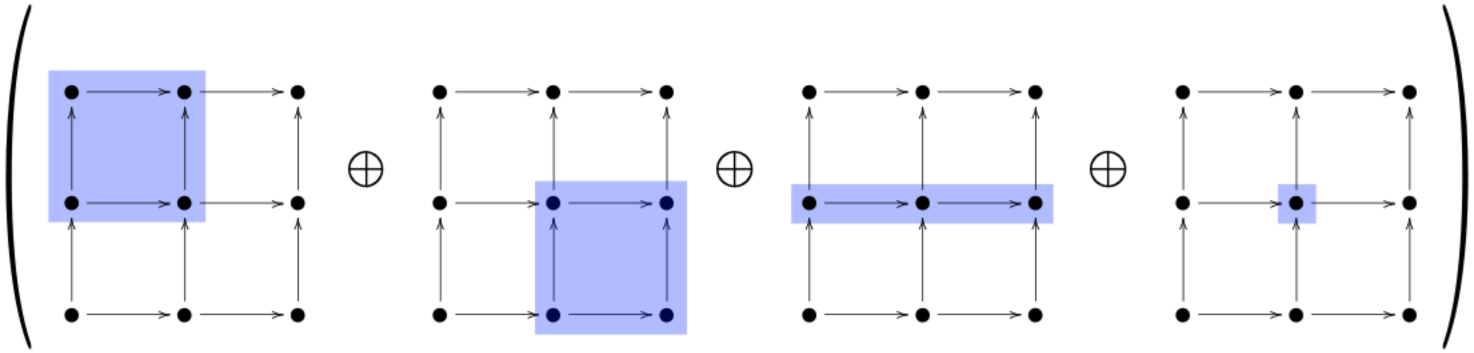
Rank decomposition

$$Rk\mathbb{V} = Rk\boldsymbol{k}_{\mathcal{R}} - Rk\boldsymbol{k}_{\mathcal{S}} = \bigoplus_{R\subset\mathcal{R}} \boldsymbol{k}_R \; - \; \bigoplus_{S\subset\mathcal{S}} \boldsymbol{k}_S$$

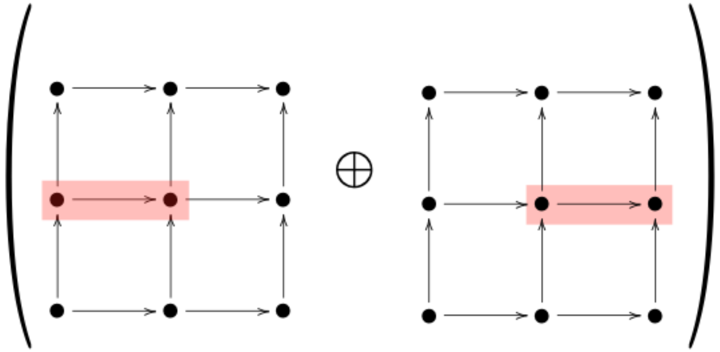
where $\mathcal{R}$ and $\mathcal{S}$ are multi-sets of rectangles

It is a minimal rank decomposition if $\mathcal{R}$ and $\mathcal{S}$ are disjoint as multi-sets.

$$\mathrm{Rk}\left(\begin{tikzcd}
\mathbf{k} \arrow[r, "\mathrm{id}"] & \mathbf{k} \arrow[r] & 0 \\
\mathbf{k} \arrow[u, "\mathrm{id}"] \arrow[r, "{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]}"] & \mathbf{k}^2 \arrow[u, "{[1\ 0]}"'] \arrow[r, "{[1\ 1]}"] & \mathbf{k} \arrow[u] \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u, "{\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]}"'] \arrow[r, "\mathrm{id}"] & \mathbf{k} \arrow[u, "\mathrm{id}"']
\end{tikzcd}\right) = \mathrm{Rk}$$

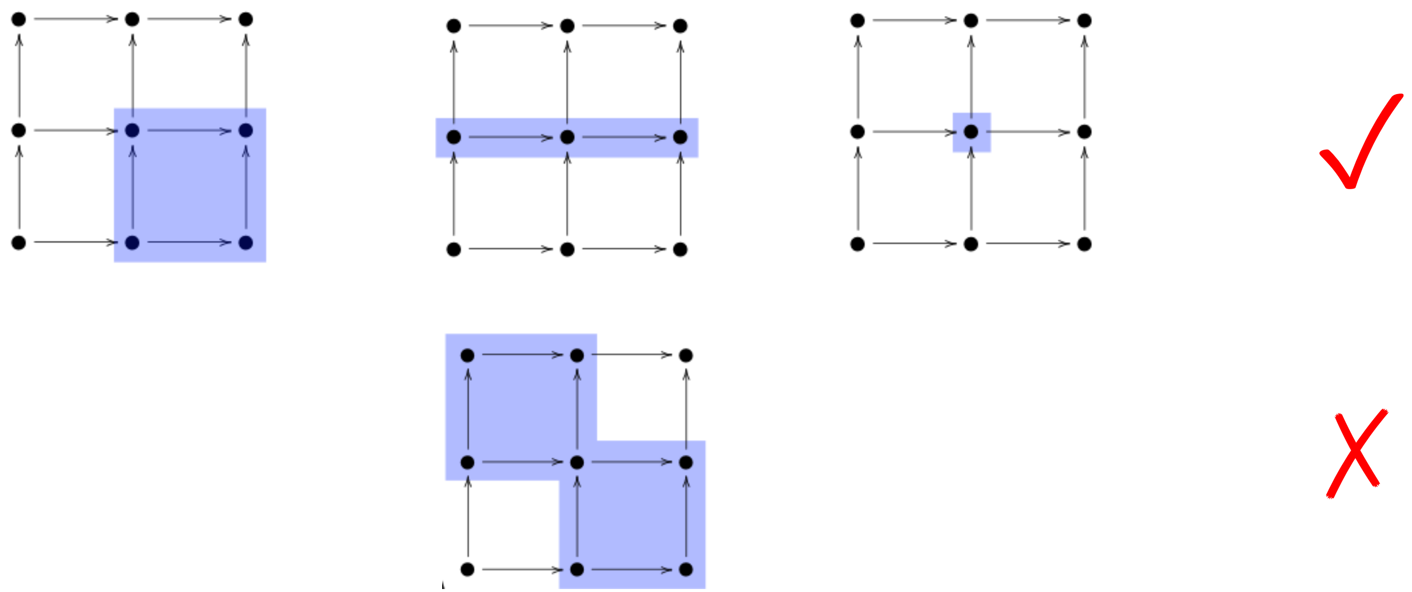


$$- \ \mathrm{Rk}$$



Rectangles

Definition 2.16 (Rectangle). *A set $R \subset \mathbb{R}^2$ is a rectangle if $R = I \times J$ where I and J are intervals in $\mathbb{R}$.*



Rectangle module

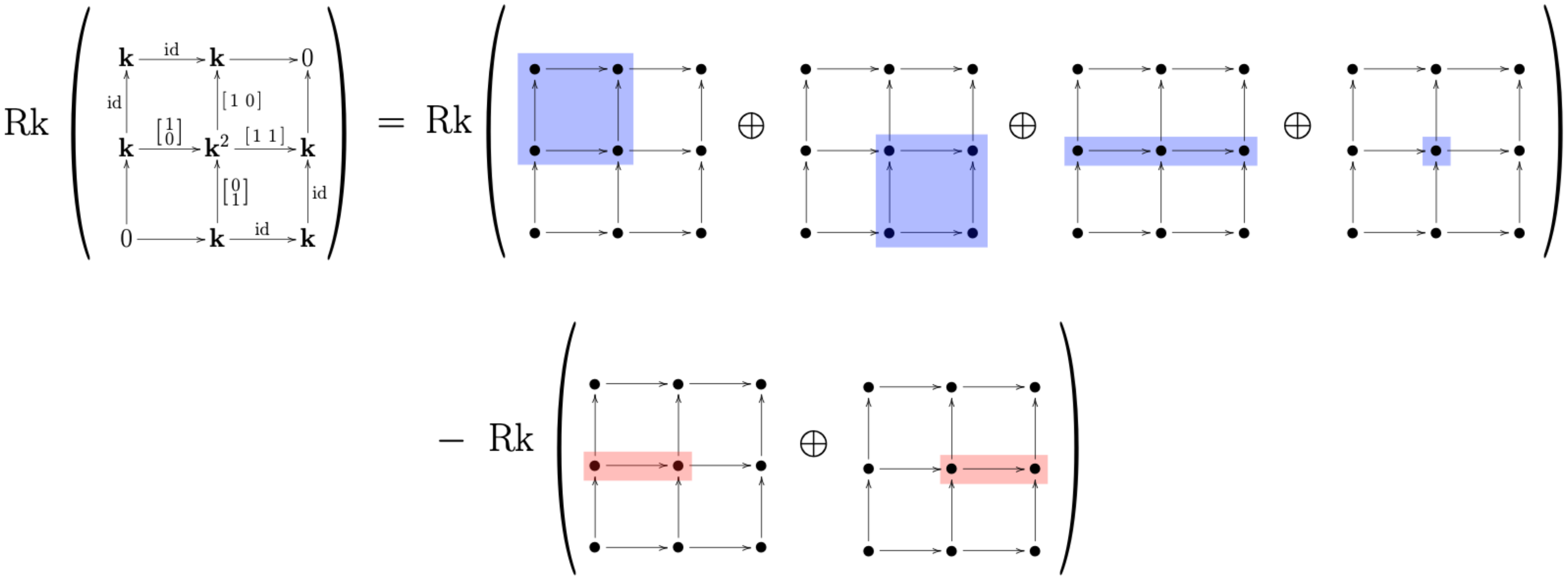
Definition 2.13 (Rectangle module). *For a rectangle $\mathcal{Q} \subset \mathbb{R}^2$, the rectangle module $k_{\mathcal{Q}}$ is defined by*

$$k_{Q,t} = \begin{cases} \boldsymbol{k}, & (t \in \mathcal{Q}) \\ 0, & (t \notin \mathcal{Q}) \end{cases} \tag{5}$$

$$k_Q(s \leq t) = \begin{cases} Id_{\boldsymbol{k}}, & (s \text{ and } t \in \mathcal{Q}) \\ 0, & else \end{cases} \tag{6}$$

Decomposition theorem

Theorem 2.3. *Given a bimodule $\mathbb{V}$ over $\mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^2$ and its rank invariant $r = \mathrm{Rk}\mathbb{V}$, the minimal rank decomposition of r is unique when it exists. Moreover, it exists whenever $\mathcal{X} \times \mathcal{Y}$ is finite or $\mathbb{V}$ is finitely presented. The multi-set of rectangles in the decomposition is called signed (persistence) barcode, denoted as Dgm $\mathbb{V}$.*



$$m((0,1),(1,2)) = 1$$
$$m((0,1),(2,1)) = 1$$
$$m((1,0),(2,1)) = 1$$
$$m((1,1),(1,1)) = 1$$
$$m((0,1),(1,1)) = -1$$
$$m((1,1),(2,1)) = -1$$

Inclusion-exclusion formula

Similar to single-parameter persistence, when $\mathcal{X} \times \mathcal{Y} \subset \mathbb{Z}^2$, the barcode Dgm$\mathbb{V}$ is easily obtained from $Rk\mathbb{V}$ via the following inclusion-exclusion formula [BLO22]:

$$\begin{aligned}
m(s,t) =& r(s,t) - r((s_x-1,s_y),t) \\
& - r((s_x,s_y-1),t) + r((s_x-1,s_y-1),t), \\
& - r(s,(t_x+1,t_y)) + r((s_x-1,s_y),(t_x+1,t_y)) \\
& + r((s_x,s_y-1),(t_x+1,t_y)) - r((s_x-1,s_y-1),(t_x+1,t_y)) \\
& - r(s,(t_x,t_y+1)) + r((s_x-1,s_y),(t_x,t_y+1)) \\
& + r((s_x,s_y-1),(t_x,t_y+1)) - r((s_x-1,s_y-1),(t_x,t_y+1)) \\
& + r(s,(t_x+1,t_y+1)) - r((s_x-1,s_y),(t_x+1,t_y+1)) \\
& - r((s_x,s_y-1),(t_x+1,t_y+1)) + r((s_x-1,s_y-1),(t_x+1,t_y+1))
\end{aligned} \tag{7}$$

Outline

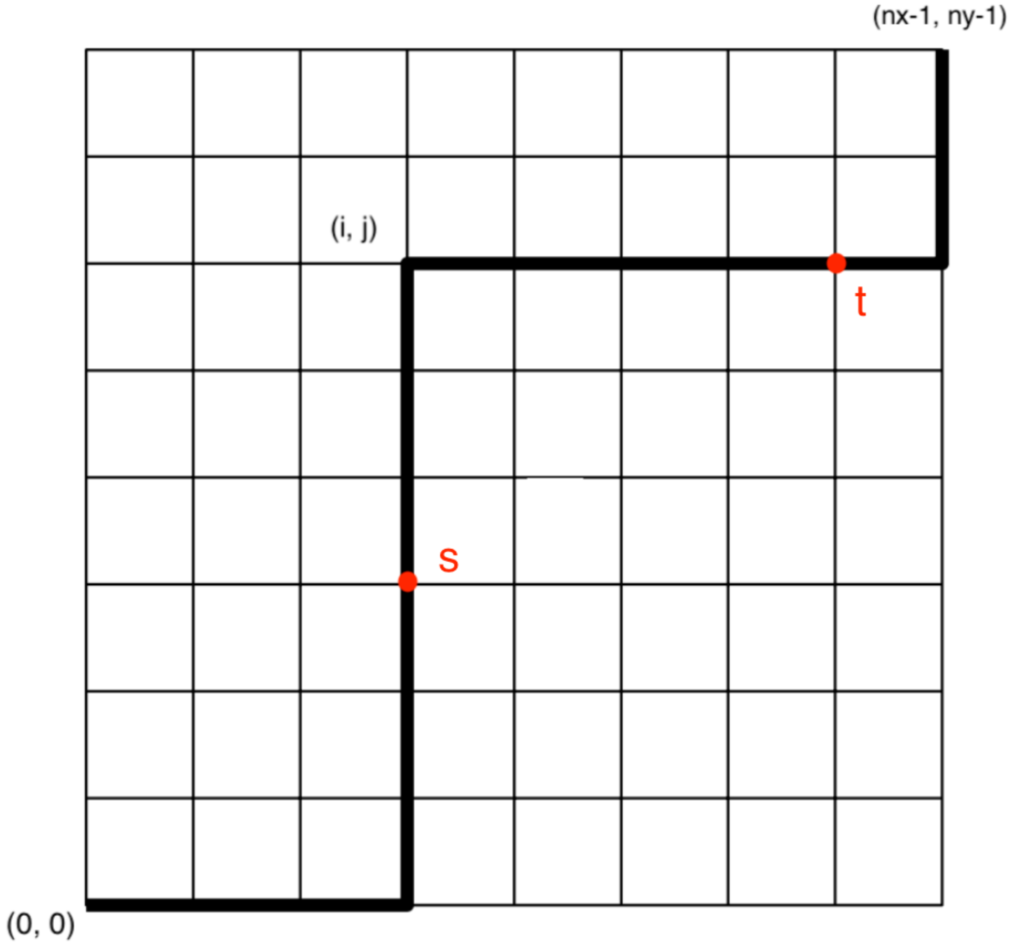
- Introduction
- Background
  - Single-parameter persistence
  - Bipersistence
- Algorithm
  - Description
  - Implementation
  - Complexity
- Experimental results
  - Correctness validation
  - Running time

Input and Output

- Input
  - A grid $G = [0, n_x - 1] \times [0, n_y - 1]$
  - A simplicial bifiltration $\mathcal{F}$ with n simplices over the grid G
  - $\mathcal{X}$ coordinate is vertical
- Output
  - Rank invariant r of the bimodule induced by $\mathcal{F}$
  - $r\colon \mathbb{Z}^2 \to \mathbb{Z}$

    $(s, t) \longmapsto r(s, t)$

Description



$For\ \forall(i,j) \in [0, n_x - 1]\times[0, n_y - 1]$

A stair $S_{i,j}$: $(0,0) \to \cdots \to (0,j) \to \cdots \to (i,j) \to \cdots \to (i, n_y - 1) \to \cdots \to (n_x - 1, n_y - 1)$

r(s, t)= #intervals in the persistence barcode passing through s and t.

Psudocode of the algorithm

- Build a simplex tree for the simplicial complex
- For each elbow/stair:
  - Update the filtration values of the simplex tree
  - Compute the persistence barcode of the single-parameter persistence
  - Calculate the rank invariant
- Calculate the barcode decompsitoin from the rank invariant

Simplex tree

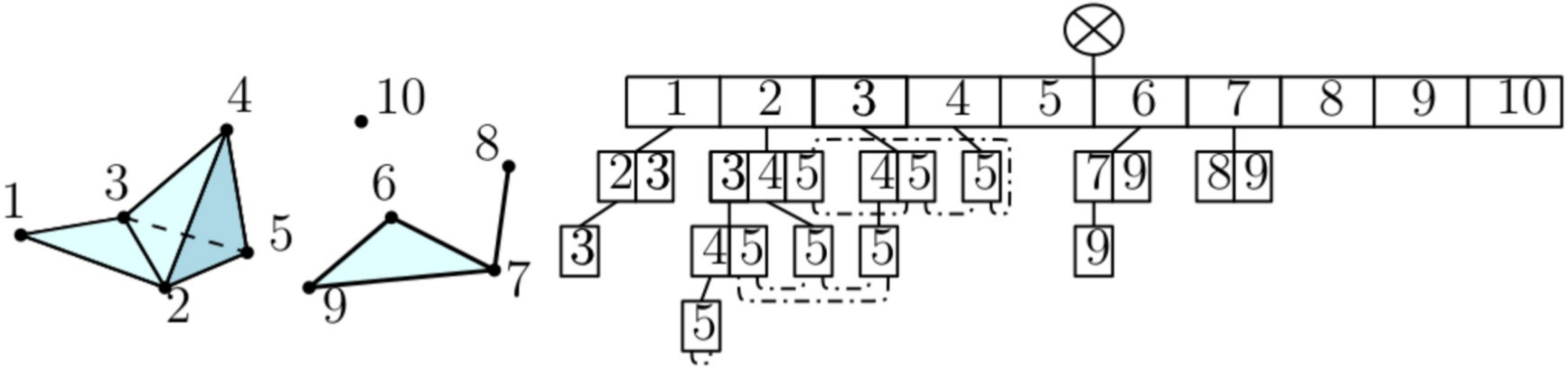
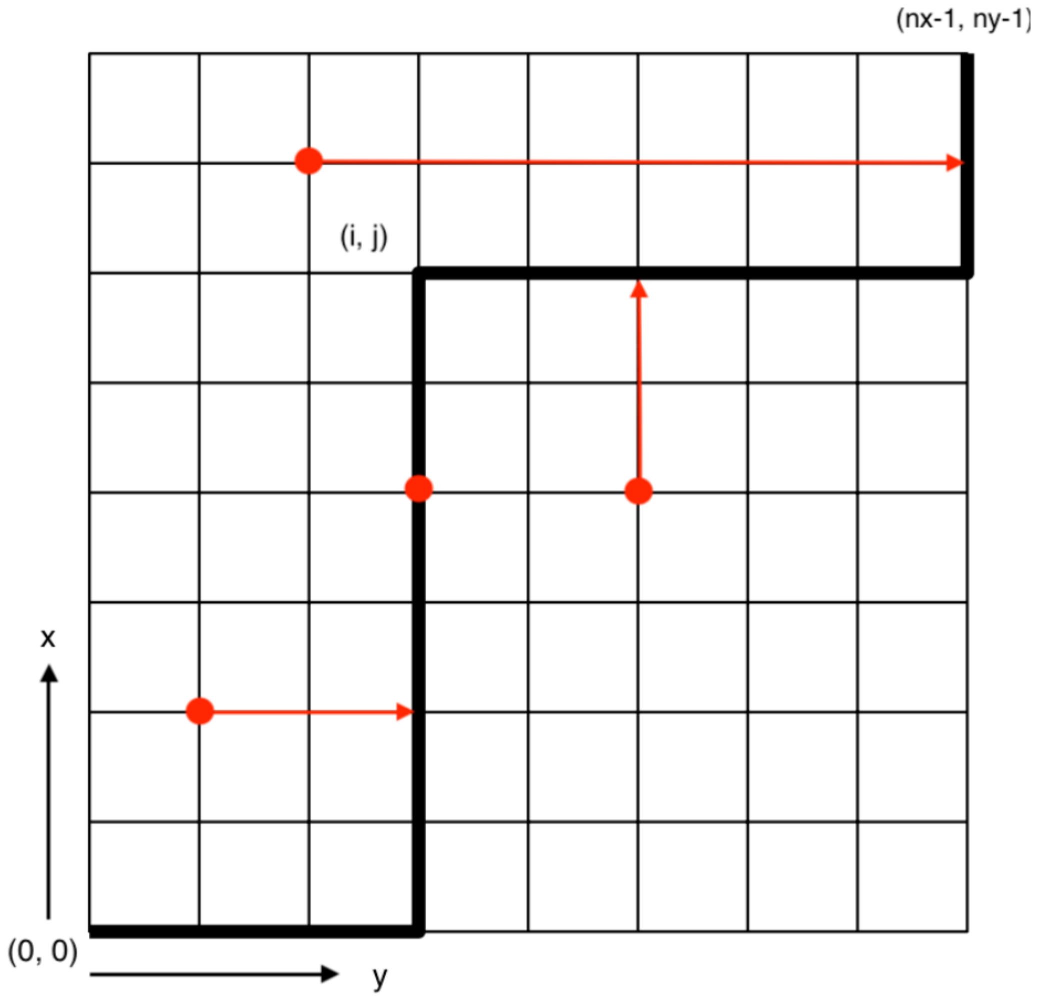


Figure 3.2: Simplex tree representation, from [The15].

Insertion and projection



The projection (x,y) *of the point* (x',y') *onto the path* $S_{(i,j)}$ *is given by:*

$$(x,y) = \begin{cases} (x',y'), & (x',y') \in S_{(i,j))} \\ (x',j), & (x',y') \notin S_{(i,j)} \text{ and } x' \leq i \text{ and } y' \leq j \\ (i,y'), & (x',y') \notin S_{(i,j)} \text{ and } x' \leq i \text{ and } y' \geq j \\ (x',n_y-1), & \text{otherwise} \end{cases}$$

Complexity

- Time
  - Time of computing 1-d persistence and updating the rank invariant is $O(n^{\omega})$ per stair
  - Time of updating the filtration values is $O(n)$ per stair
  - $n_x n_y$ stairs,
  - $O(n_x n_y n^{\omega})$ in total. In practice, we expect $O\left(n_x n_y n\right)$
- Space
  - $O(nm)$
  - n is the number of simplices
  - m is the maximum dimension of faces in the simplicial complex

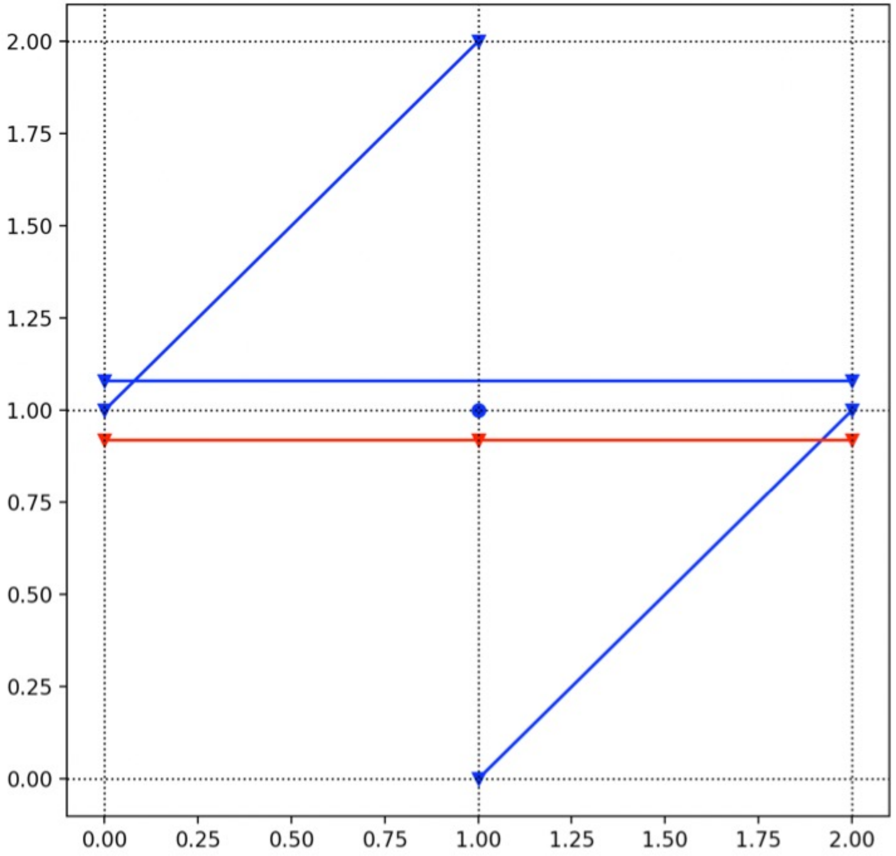
Outline

$$m((0,1),(1,2)) = 1$$
$$m((0,1),(2,1)) = 1$$
$$m((1,0),(2,1)) = 1$$
$$m((1,1),(1,1)) = 1$$
$$m((0,1),(1,1)) = -1$$
$$m((1,1),(2,1)) = -1$$

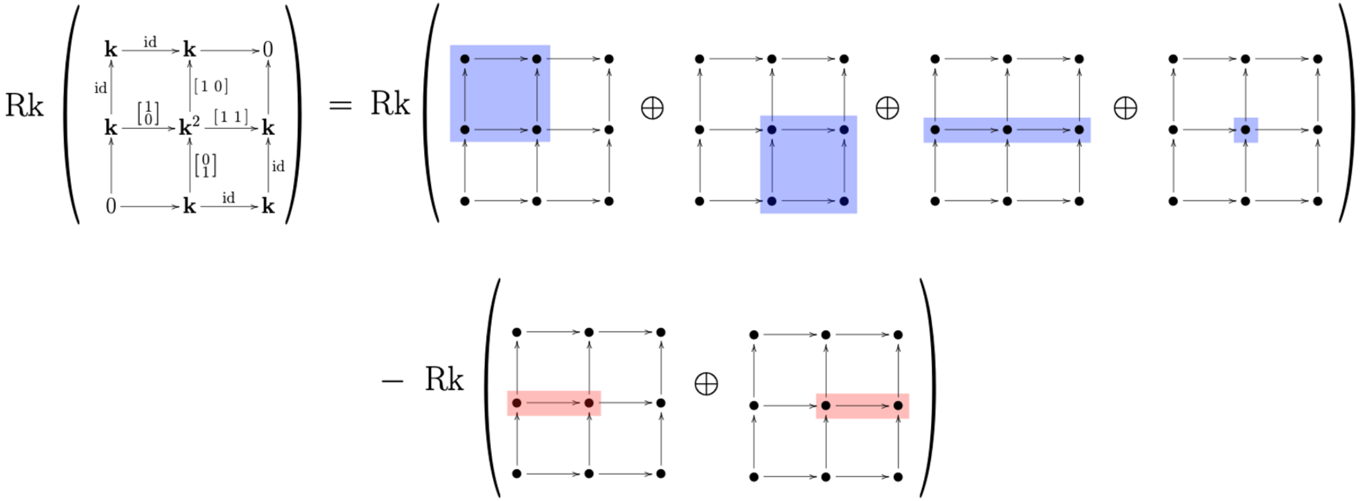
A toy example

$$\begin{tikzcd}
\mathbf{k} \arrow[r, "\mathrm{id}"] & \mathbf{k} \arrow[r] & 0 \\
\mathbf{k} \arrow[u, "\mathrm{id}"] \arrow[r, "\begin{bmatrix}1\\0\end{bmatrix}"] & \mathbf{k}^2 \arrow[u, "\begin{bmatrix}1 & 0\end{bmatrix}"'] \arrow[r, "\begin{bmatrix}1 & 1\end{bmatrix}"] & \mathbf{k} \arrow[u] \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u, "\begin{bmatrix}0\\1\end{bmatrix}"'] \arrow[r, "\mathrm{id}"] & \mathbf{k} \arrow[u, "\mathrm{id}"']
\end{tikzcd}$$

(a)



A toy example

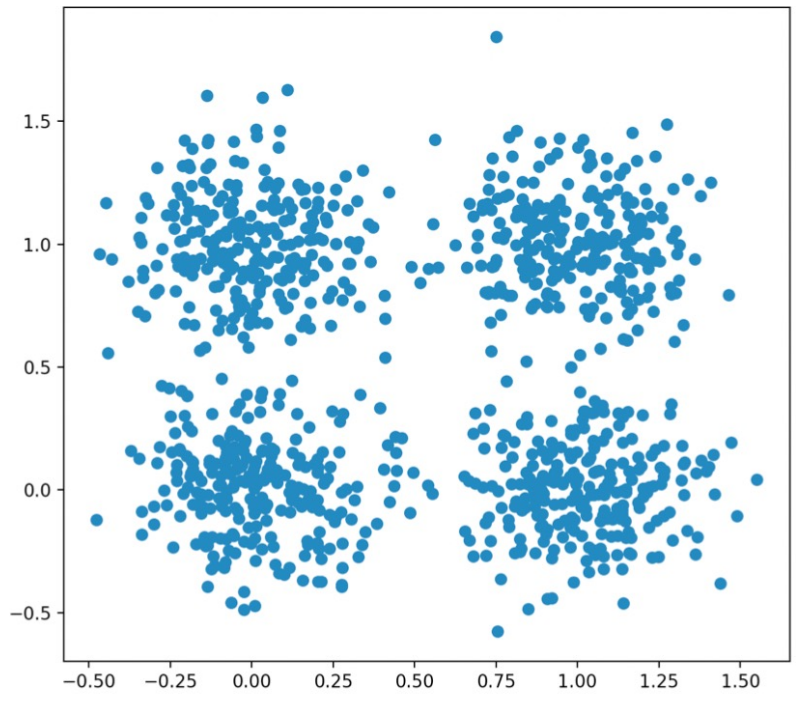


$$
Rk\left(\begin{tikzcd}
\mathbf{k} \arrow[r, "id"] & \mathbf{k} \arrow[r] & 0 \\
\mathbf{k} \arrow[u, "id"] \arrow[r, "\begin{bmatrix}1\\0\end{bmatrix}"] & \mathbf{k}^2 \arrow[u, "\begin{bmatrix}1 & 0\end{bmatrix}"'] \arrow[r, "\begin{bmatrix}1 & 1\end{bmatrix}"] & \mathbf{k} \arrow[u] \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u, "\begin{bmatrix}0\\1\end{bmatrix}"] \arrow[r, "id"] & \mathbf{k} \arrow[u, "id"']
\end{tikzcd}\right) =
$$

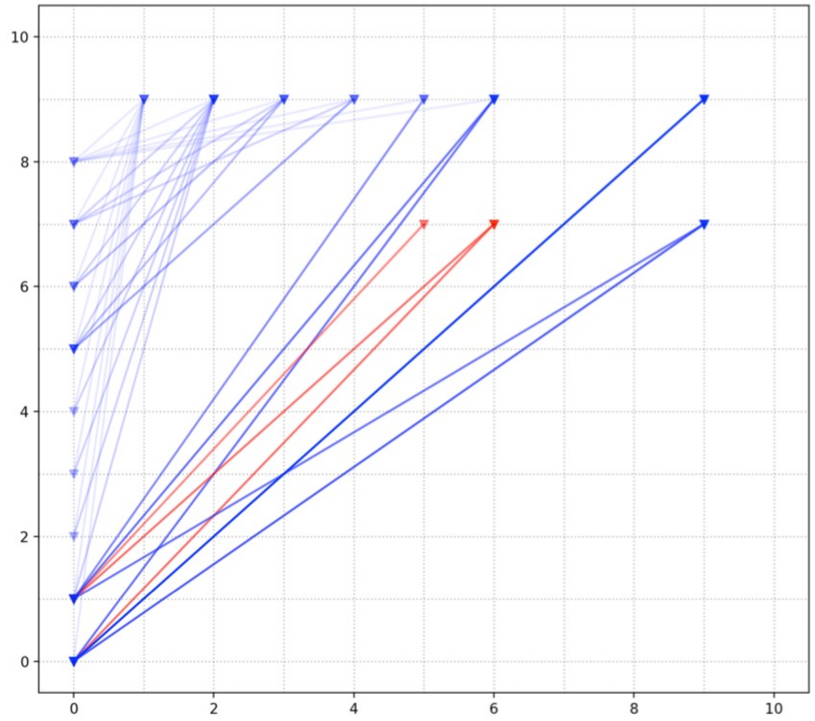
$$
Rk\left(\begin{tikzcd}
0 \arrow[r] & 0 \arrow[r] & 0 \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u] \arrow[r, "id"] & \mathbf{k} \arrow[u] \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u, "id"'] \arrow[r, "id"] & \mathbf{k} \arrow[u, "id"']
\end{tikzcd}
\oplus
\begin{tikzcd}
0 \arrow[r] & \mathbf{k} \arrow[r] & 0 \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u, "id"'] \arrow[r] & 0 \arrow[u] \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u, "id"'] \arrow[r] & 0 \arrow[u]
\end{tikzcd}
\oplus
\begin{tikzcd}
\mathbf{k} \arrow[r, "id"] & \mathbf{k} \arrow[r] & 0 \\
\mathbf{k} \arrow[u, "id"'] \arrow[r, "id"] & \mathbf{k} \arrow[u, "id"'] \arrow[r] & 0 \arrow[u] \\
0 \arrow[u] \arrow[r] & 0 \arrow[u] \arrow[r] & 0 \arrow[u]
\end{tikzcd}
\oplus
\begin{tikzcd}
0 \arrow[r] & 0 \arrow[r] & 0 \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u] \arrow[r] & 0 \arrow[u] \\
0 \arrow[u] \arrow[r] & 0 \arrow[u] \arrow[r] & 0 \arrow[u]
\end{tikzcd}\right)
$$

$$
-Rk\left(\begin{tikzcd}
0 \arrow[r] & 0 \arrow[r] & 0 \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u] \arrow[r] & 0 \arrow[u] \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u, "id"'] \arrow[r] & 0 \arrow[u]
\end{tikzcd}
\oplus
\begin{tikzcd}
0 \arrow[r] & \mathbf{k} \arrow[r] & 0 \\
0 \arrow[u] \arrow[r] & \mathbf{k} \arrow[u, "id"'] \arrow[r] & 0 \arrow[u] \\
0 \arrow[u] \arrow[r] & 0 \arrow[u] \arrow[r] & 0 \arrow[u]
\end{tikzcd}\right)
$$

Bimodule clustering



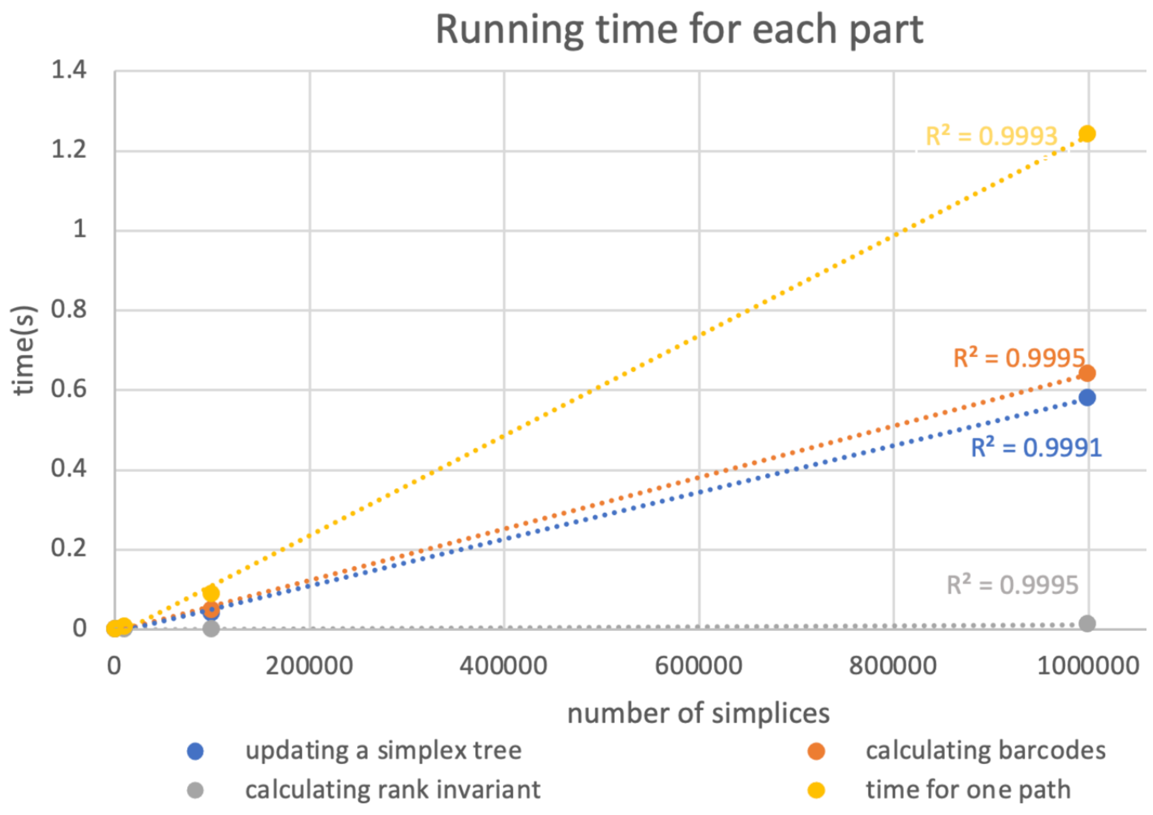
(a)



(b)

Figure 4.2: (a): Point cloud with 1000 points sampled from 4 Gaussian distributions (b): The corresponding signed barcode over a 10×10 grid. Each bar with endpoints $s \leq t$ in the barcode has an intensity proportional to $\min\{t_x - s_x, t_y - s_y\}$; in particular, horizontal and vertical bars are invisible

Running time



PART	SIMPLICES	GRID SIZE	TIME(S)
UPDATING A SIMPLEX TREE	1025	20×20	0.00022
	10007	20×20	0.0025
	100000	20×20	0.039
	1000007	20×20	0.58
CALCULATING BARCODES	1025	20×20	0.00032
	10007	20×20	0.0036
	100000	20×20	0.048
	1000007	20×20	0.64
CALCULATING RANK INVARIANT	1025	20×20	9.6E-05
	10007	20×20	0.00015
	100000	20×20	0.001
	1000007	20×20	0.012
TIME FOR ONE PATH	1025	20×20	0.00076
	10007	20×20	0.0064
	100000	20×20	0.088
	1000007	20×20	1.24

Complexity

- Time
  - Time of computing 1-d persistence and updating the rank invariant is $O(n^{\omega})$ per stair
  - Time of updating the filtration values is $O(n)$ per stair
  - $n_x n_y$ stairs,
  - $O(n_x n_y n^{\omega})$ in total. In practice, we expect $O\left(n_x n_y n\right)$
- Space
  - $O(nm)$
  - n is the number of simplices
  - m is the maximum dimension of faces in the simplicial complex

Running time

PART	SIMPLICES	GRID SIZE	TIME(S)
BARCODE DECOMPOSITION	1025	20×20	0.00038
	10007	20×20	0.00054
	100000	20×20	0.00045
	1000007	20×20	0.00054
TOTAL TIME	1025	20×20	0.31
	10007	20×20	2.58
	100000	20×20	35.32
	1000007	20×20	495.63

Conclusion

- Introduce a new algorithm for rank invariant.
- Decent running time in practice. There exists one algorithm whose theoretical time complexity $O(n^\omega + n_x n_y n)$ does match with the practical time complexity we observed. With that time complexity, their algorithm is much more complicated than ours, so we already have the best possible running time in practice.
- Future work: higher dimension, degree 0 homology case…

Thank you!